



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/49, C07K 14/16		A2	(11) International Publication Number: WO 00/39303 (43) International Publication Date: 6 July 2000 (06.07.00)
(21) International Application Number: PCT/US99/31272		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, ER, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 30 December 1999 (30.12.99)			
(30) Priority Data: 60/114,495 31 December 1998 (31.12.98) US 60/156,670 29 September 1999 (29.09.99) US			
(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).			
(72) Inventors: BARNETT, Susan; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). HARTOG, Karin; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). MARTIN, Eric; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US).			
(74) Agents: DOLLARD, Anne, S.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US) et al.			
(54) Title: MODIFIED HIV ENV POLYPEPTIDES			
(57) Abstract			
Polynucleotide encoding modified HIV Env polypeptides are disclosed. The Env polypeptides are modified so as to expose at least part of the CD4 binding region. Methods of diagnosis, treatment and prevention using the polynucleotides and polypeptides are also provided.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Bolivia	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BF	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Bosnia	IK	Iceland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

MODIFIED HIV ENV POLYPEPTIDESTechnical Field

5 The invention relates generally to modified HIV envelope (Env) polypeptides which are useful as immunizing agents or for generating an immune response in a subject, for example a cellular immune response or a protective immune response. More particularly, the invention relates Env polypeptides such as gp120, gp140 or gp160, wherein at least one of the native β -sheet configurations has been modified. The invention also pertains to methods 10 of using these polypeptides to elicit an immune response against a broad range of HIV subtypes.

Background of the Invention

15 The human immunodeficiency virus (HIV-1, also referred to as HTLV-III, LAV or HTLV-III/LAV) is the etiological agent of the acquired immune deficiency syndrome (AIDS) and related disorders. (see, e.g., Barre-Sinoussi, et al., (1983) *Science* 220:868-871; Gallo et al. (1984) *Science* 224:500-503; Levy et al., (1984) *Science* 225:840-842; Siegal et al., (1981) *N. Engl. J. Med.* 305:1439-1444). AIDS patients usually have a long asymptomatic period followed by the progressive degeneration of the immune system and the central nervous 20 system. Replication of the virus is highly regulated, and both latent and lytic infection of the CD4 positive helper subset of T-lymphocytes occur in tissue culture (Zagury et al., (1986) *Science* 231:850-853). Molecular studies of HIV-1 show that it encodes a number of genes (Ratner et al., (1985) *Nature* 313:277-284; Sanchez-Pescador et al., (1985) *Science* 227:484-492), including three structural genes -- gag, pol and env -- that are common to all 25 retroviruses. Nucleotide sequences from viral genomes of other retroviruses, particularly HIV-2 and simian immunodeficiency viruses, SIV (previously referred to as STLV-III), also contain these structural genes. (Guyader et al., (1987) *Nature* 326:662-669; Chakrabarti et al., (1987) *Nature*

30 The envelope protein of HIV-1, HIV-2 and SIV is a glycoprotein of about 160 kd (gp160). During virus infection of the host cell, gp160 is cleaved by host cell proteases to form gp120 and the integral membrane protein, gp41. The gp41 portion is anchored in the

membrane bilayer of virion, while the gp120 segment protrudes into the surrounding environment. gp120 and gp41 are more covalently associated and free gp120 can be released from the surface of virions and infected cells.

As depicted in Figure 1, crystallography studies of the gp120 core polypeptide 5 indicate that this polypeptide is folded into two major domains having certain emanating structures. The inner domain (inner with respect to the N and C terminus) features a two-helix, two-stranded bundle with a small five-stranded β -sandwich at its termini-proximal end and a projection at the distal end from which the V1/V2 stem emanates. The outer domain is a staked double barrel that lies along side the inner domain so that the outer barrel and inner 10 bundle axes are approximately parallel. Between the distal inner domain and the distal outer domain is a four-stranded bridging sheet which holds a peculiar minidomain in contact with, but distinct from, the inner, the outer domain, and the V1/V2 domain. The bridging sheet is composed of four β -strand structures (β -3, β -2, β -21, β -20, shown in Figure 1). The bridging region can be seen in Figure 1 packing primarily over the inner domain, although some 15 surface residues of the outer domain, such as Phe 382, reach into the bridging sheet to form part of its hydrophobic core.

The basic unit of the β -sheet conformation of the bridging sheet region is the β -strand which exists as a less tightly coiled helix, with 2.0 residues per turn. The β -strand conformation is only stable when incorporated into a β -sheet, where hydrogen bonds with 20 close to optimal geometry are formed between the peptide groups on adjacent β -strands; the dipole moments of the strands are also aligned favorably. Side chains from adjacent residues of the same strand protrude from opposite sides of the sheet and do not interact with each other, but have significant interactions with their backbone and with the side chains of neighboring strands. For a general description of β -sheets, see, e.g., T.E. Creighton, Proteins: 25 Structures and Molecular Properties (W.H. Freeman and Company, 1993); and A.L. Lehninger, Biochemistry (Worth Publishers, Inc., 1975).

The gp120 polypeptide is instrumental in mediating entry into the host cell. Recent studies have indicated that binding of CD4 to gp120 induces a conformational change in Env that allows for binding to a co-receptor (e.g. a chemokine receptor) and subsequent entry of 30 the virus into the cell. (Wyatt, R., et al. (1998) *Nature* 393:705-711; Kwong, P., et al.(1998) *Nature* 393:648-659). Referring again to Figure 1, CD4 is bound into a depression formed at the interface of the outer domain, the inner domain and the bridging sheet of gp120.

Immunogenicity of the gp120 polypeptide has also been studied. For example, individuals infected by HIV-1 usually develop antibodies that can neutralize the virus in *in vitro* assays, and this response is directed primarily against linear neutralizing determinants in the third variable loop of gp120 glycoprotein (Javaherian, K., et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6786-6772; Matsushita, M., et al. (1988) *J. Virol.* 62:2107-2144; Putney, S., et al. (1986) *Science* 234:1392-1395; Rushe, J. R., et al. (1988) *Proc. Natl. Acad. Sci. USA* 85: 3198-3202.). However, these antibodies generally exhibit the ability to neutralize only a limited number of HIV-1 strains (Matthews, T. (1986) *Proc. Natl. Acad. Sci. USA* 83:9709-9713; Nara, P. L., et al. (1988) *J. Virol.* 62:2622-2628; Parker, T. J., et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:1932-1936). Later in the course of HIV infection in humans, antibodies capable of neutralizing a wider range of HIV-1 isolates appear (Barre-Sinoussi, F., et al. (1983) *Science* 220:868-871; Robert-Guroff, M., et al. (1985) *Nature* (London) 316:72-74; Weis, R., et al. (1985) *Nature* (London) 316:69-72; Weis, R., et al. (1986) *Nature* (London) 324:572-575).

Recent work done by Stamatatos et al (1998) *AIDS Res Hum Retroviruses* 14(13):1129-39, shows that a deletion of the variable region 2 from a HIV-1_{SF162} virus, which utilizes the CCR-5 co-receptor for virus entry, rendered the virus highly susceptible to serum-mediated neutralization. This V2 deleted virus was also neutralized by sera obtained from patients infected not only with clade B HIV-1 isolates but also with clade A, C, D and F HIV-1 isolates. However, deletion of the variable region 1 had no effect. Deletion of the variable regions 1 and 2 from a LAI isolate HIV-1_{BB} also increased the susceptibility to neutralization by monoclonal antibodies whose epitopes are located within the V3 loop, the CD4-binding site, and conserved gp120 regions (Wyatt, R., et al. (1995) *J Virol.* 69:5723-5733). Rabbit immunogenicity studies done with the HIV-1 virus with deletions in the V1/V2 and V3 region from the LAI strain, which uses the CXCR4 co-receptor for virus entry, showed no improvement in the ability of Env to raise neutralizing antibodies (Leu et al. (1998) *AIDS Res. and Human Retroviruses*. 14:151-155).

Further, a subset of the broadly reactive antibodies, found in most infected individuals, interferes with the binding of gp120 and CD4 (Kang, C.-Y., et al. (1991) *Proc. Natl. Acad. Sci. USA*. 88:6171-6175; McDougal, J. S., et al. (1986) *J. Immunol.* 137:2937-2944). Other antibodies are believed to bind to the chemokine receptor binding region after CD4 has bound to Env (Thali et al. (1993) *J. Virol.* 67:3978-3988). The fact that neutralizing

antibodies generated during the course of HIV infection do not provide permanent antiviral effect may in part be due to the generation of "neutralization escapes" virus mutants and to the general decline in the host immune system associated with pathogenesis. In contrast, the presence of pre-existing neutralizing antibodies upon initial HIV-1 exposure will likely have

5 a protective effect.

It is widely thought that a successful vaccine should be able to induce a strong, broadly neutralizing antibody response against diverse HIV-1 strains (Montefiori and Evans (1999) *AIDS Res. Hum. Ret.* 15(8):689-698; Bolognesi, D.P., et al. (1994) *Ann. Int. Med.* 8:603-611; Haynes, B., F., et al. (1996) *Science* ;271: 324-328.). Neutralizing antibodies, by

10 attaching to the incoming virions, can reduce or even prevent their infectivity for target cells and prevent the cell-to-cell spread of virus in tissue culture (Hu et al. (1992) *Science* 255:456-459; Burton, D.,R. and Montefiori, D. (1997) *AIDS* 11(suppl. A): 587-598). However as described above, antibodies directed against gp120 do not generally exhibit broad antibody responses against different HIV strains.

15 Currently, the focus of vaccine development, from the perspective of humoral immunity, is on the neutralization of primary isolates that utilize the CCR5 chemokine co-receptor believed to be important in virus entry (Zhu, T., et al. (1993) *Science* 261:1179-1181; Fiore, J., et al. (1994) *Virology*; 204:297-303). These viruses are generally much more resistant to antibody neutralization than T-cell line adapted strains that use the CXCR4 co-

20 receptor, although both can be neutralized *in vitro* by certain broadly and potent acting monoclonal antibodies, such as IgG1b12, 2G12 and 2F5 (Trkola, A., et al. (1995) *J. Virol.* 69:6609-6617; D'Sousa PM., et al (1997) *J. Infect. Dis.* 175:1062-1075). These monoclonal antibodies are directed to the CD4 binding site, a glycosylation site and to the gp41 fusion domain, respectively. The problem that remains, however, is that it is not known how to

25 induce antibodies of the appropriate specificity by vaccination. Antibodies (Abs) elicited by gp120 glycoprotein from a given isolate are usually only able to neutralize closely related viruses generally from similar, usually from the same, HIV-1 subtype.

30 Despite the above approaches, there remains a need for Env antigens that can elicit an immunological response (e.g., neutralizing and/or protective antibodies) in a subject against multiple HIV strains and subtypes, for example when administered as a vaccine. The present invention solves these and other problems by providing modified Env polypeptides (e.g., gp120) to expose epitopes in or near the CD4 binding site.

Summary of the Invention

In accordance with the present invention, modified HIV Env polypeptides are provided. In particular, deletions and/or mutations are made in one or more of the 4- β antiparallel-bridging sheet in the HIV Env polypeptide. In this way, enough structure is left 5 to allow correct folding of the polypeptide, for example of gp120, yet enough of the bridging sheet is removed to expose the CD4 groove, allowing an immune response to be generated against epitopes in or near the CD4 binding site of the Env polypeptide (e.g., gp120).

In one aspect, the invention includes a polynucleotide encoding a modified HIV Env polypeptide wherein the polypeptide has at least one modified (e.g., deleted or replaced) 10 amino acid residue deleted in the region corresponding to residues 421 to 436 relative to HXB-2, for example the constructs depicted in Figures 6-29 (SEQ ID NOs:3 to 26). In certain embodiments, the polynucleotide also has the region corresponding to residues 124-198 of the polypeptide HXB-2 (e.g., V1/V2) deleted and at least one amino acid deleted or replaced in the regions corresponding to the residues 119 to 123 and 199 to 210, relative to 15 HXB-2. In other embodiments, these polynucleotides encode Env polypeptides having at least one amino acid of the small loop of the bridging sheet (e.g., amino acid residues 427 to 429 relative to HXB-2) deleted or replaced. The amino acid sequences of the modified polypeptides encoded by the polynucleotides of the present invention can be based on any HIV variant, for example SF162.

20 In another aspect, the invention includes immunogenic modified HIV Env polypeptides having at least one modified (e.g., deleted or replaced) amino acid residue deleted in the region corresponding to residues 421 to 436 relative to HXB-2, for example a deletion or replacement of one amino acids in the small loop region (e.g., amino acid residues 427 to 429 relative to HXB-2). These polypeptides may have modifications (e.g., a deletion 25 or a replacement) of at least one amino acid between about amino acid residue 420 and amino acid residue 436, relative to HXB-2 and, optionally, may have deletions or truncations of the V1 and/or V2 regions. The immunogenic, modified polypeptides of the present invention can be based on any HIV variant, for example SF162.

In another aspect, the invention includes a vaccine composition comprising any of the 30 polynucleotides encoding modified Env polypeptides described above. Vaccine compositions comprising the modified Env polypeptides and, optionally, an adjuvant are also included in the invention.

In yet another aspect, the invention includes a method of inducing an immune response in subject comprising, administering one or more of the polynucleotides or constructs described above in an amount sufficient to induce an immune response in the subject. In certain embodiments, the method further comprises administering an adjuvant to 5 the subject.

In another aspect, the invention includes a method of inducing an immune response in a subject comprising administering a composition comprising any of the modified Env polypeptides described above and an adjuvant. The composition is administered in an amount sufficient to induce an immune response in the subject.

10 In another aspect, the invention includes a method of inducing an immune response in a subject comprising

- (a) administering a first composition comprising any of the polynucleotides described above in a priming step and
- (b) administering a second composition comprising any of the modified Env

15 polypeptides described above, as a booster, in an amount sufficient to induce an immune response in the subject. In certain embodiments, the first composition, the second composition or both the first and second compositions further comprise an adjuvant.

These and other embodiments of the subject invention will readily occur to those of skill in the art in light of the disclosure herein.

20

Brief Description of the Drawings

Figure 1 is a schematic depiction of the tertiary structure of the HIV-1_{HXB-2} Env gp120 polypeptide, as determined by crystallography studies.

25 Figures 2A-C depict alignment of the amino acid sequence of wild-type HIV-1_{HXB-2} Env gp160 polypeptide (SEQ ID NO:1) with amino acid sequence of HIV variants SF162 (shown as "162") (SEQ ID NO:2), SF2, CM236 and US4. Arrows indicate the regions that are deleted or replaced in the modified polypeptides. Black dots indicate conserved cysteine residues. The star indicates the position of the last amino acid in gp120.

30 Figures 3A-J depict alignment of nucleotide sequences of polynucleotides encoding modified Env polypeptides having V1/V2 deletions. The unmodified amino acid residues encoded by these sequences correspond to wildtype SF162 residues but are numbered relative to HXB-2.

Figures 4A-M depict alignment of nucleotide sequences of polynucleotides encoding modified Env polypeptides having deletions or replacements in the small loop. The unmodified amino acid residues encoded by these sequences correspond to wildtype SF162 residues but are numbered relative to HXB-2.

5 Figures 5A-N depict alignment of nucleotide sequences of polynucleotides encoding modified Env polypeptides having both V1/V2 deletions and, in addition, deletions or replacements in the small loop. The unmodified amino acid residues encoded by these sequences correspond to wildtype SF162 residues but are numbered relative to HXB-2.

Figure 6 depicts the nucleotide sequence of the construct designated Val120-Ala204
10 (SEQ ID NO:3).

Figure 7 depicts the nucleotide sequence of the construct designated Val120-Ile201
(SEQ ID NO:4).

Figure 8 depicts the nucleotide sequence of the construct designated Val120-Ile201B
(SEQ ID NO:5).

15 Figure 9 depicts the nucleotide sequence of the construct designated Lys121-Val200
(SEQ ID NO:6).

Figure 10 depicts the nucleotide sequence of the construct designated Leu122-Ser199
(SEQ ID NO:7).

Figure 11 depicts the nucleotide sequence of the construct designated Val120-Thr202
20 (SEQ ID NO:8).

Figure 12 depicts the nucleotide sequence of the construct designated Trp427-Gly431
(SEQ ID NO:9).

Figure 13 depicts the nucleotide sequence of the construct designated Arg426-Gly431
(SEQ ID NO:10).

25 Figure 14 depicts the nucleotide sequence of the construct designated Arg426-
Gly431B (SEQ ID NO:11).

Figure 15 depicts the nucleotide sequence of the construct designated Arg426-Lys432
(SEQ ID NO:12).

Figure 16 depicts the nucleotide sequence of the construct designated Asn425-Lys432
30 (SEQ ID NO:13).

Figure 17 depicts the nucleotide sequence of the construct designated Ile424-Ala433
(SEQ ID NO:14).

Figure 18 depicts the nucleotide sequence of the construct designated Ile423-Met434 (SEQ ID NO:15).

Figure 19 depicts the nucleotide sequence of the construct designated Gln422-Tyr435 (SEQ ID NO:16).

5 Figure 20 depicts the nucleotide sequence of the construct designated Gln422-Tyr435B (SEQ ID NO:17).

Figure 21 depicts the nucleotide sequence of the construct designated Leu122-Ser199;Arg426-Gly431 (SEQ ID NO:18).

10 Figure 22 depicts the nucleotide sequence of the construct designated Leu122-Ser199;Arg426-Lys432 (SEQ ID NO:19).

Figure 23 depicts the nucleotide sequence of the construct designated Leu122-Ser199;Trp427-Gly431 (SEQ ID NO:20).

Figure 24 depicts the nucleotide sequence of the construct designated Lys121-Val200;Asn425-Lys432 (SEQ ID NO:21).

15 Figure 25 depicts the nucleotide sequence of the construct designated Val120-Ile201;Ile424-Ala433 (SEQ ID NO:22).

Figure 26 depicts the nucleotide sequence of the construct designated Val120-Ile201B; Ile424-Ala433 (SEQ ID NO:23).

20 Figure 27 depicts the nucleotide sequence of the construct designated Val120-Thr202;Ile424-Ala433 (SEQ ID NO:24).

Figure 28 depicts the nucleotide sequence of the construct designated Val127-Asn195 (SEQ ID NO:25).

Figure 29 depicts the nucleotide sequence of the construct designated Val127-Asn195; Arg426-Gly431 (SEQ ID NO:26).

25

Detailed Description of the Invention

The practice of the present invention will employ, unless otherwise indicated, conventional methods of protein chemistry, viral immunobiology, molecular biology and recombinant DNA techniques within the skill of the art. Such techniques are explained fully 30 in the literature. See, e.g., T.E. Creighton, Proteins: Structures and Molecular Properties (W.H. Freeman and Company, 1993); Nelson L.M. and Jerome H.K. HIV Protocols in Methods in Molecular Medicine, vol. 17, 1999; Sambrook, et al., Molecular Cloning: A

Laboratory Manual (Cold Spring Harbor Laboratory, 1989); F.M. Ausubel et al. Current Protocols in Molecular Biology, Greene Publishing Associates & Wiley Interscience New York; and Lipkowitz and Boyd, Reviews in Computational Chemistry, volumes 1-present (Wiley-VCH, New York, New York, 1999).

5 It must be noted that, as used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a polypeptide" includes a mixture of two or more polypeptides, and the like.

10 **Definitions**

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

15 The terms "polypeptide," and "protein" are used interchangeably herein to denote any polymer of amino acid residues. The terms encompass peptides, oligopeptides, dimers, multimers, and the like. Such polypeptides can be derived from natural sources or can be synthesized or recombinantly produced. The terms also include postexpression modifications of the polypeptide, for example, glycosylation, acetylation, phosphorylation, etc.

A polypeptide as defined herein is generally made up of the 20 natural amino acids Ala (A), Arg (R), Asn (N), Asp (D), Cys (C), Gln (Q), Glu (E), Gly (G), His (H), Ile (I), Leu (L), Lys (K), Met (M), Phe (F), Pro (P), Ser (S), Thr (T), Trp (W), Tyr (Y) and Val (V) and may also include any of the several known amino acid analogs, both naturally occurring and synthesized analogs, such as but not limited to homoisoleucine, asaleucine, 2-(methylenecyclopropyl)glycine, S-methylcysteine, S-(prop-1-enyl)cysteine, homoserine, ornithine, norleucine, norvaline, homoarginine, 3-(3-carboxyphenyl)alanine, 25 cyclohexylalanine, mimosine, pipecolic acid, 4-methylglutamic acid, canavanine, 2,3-diaminopropionic acid, and the like. Further examples of polypeptide agents which will find use in the present invention are set forth below.

30 By "geometry" or "tertiary structure" of a polypeptide or protein is meant the overall 3-D configuration of the protein. As described herein, the geometry can be determined, for example, by crystallography studies or by using various programs or algorithms which predict the geometry based on interactions between the amino acids making up the primary and secondary structures.

By "wild type" polypeptide, polypeptide agent or polypeptide drug, is meant a naturally occurring polypeptide sequence, and its corresponding secondary structure. An "isolated" or "purified" protein or polypeptide is a protein which is separate and discrete from a whole organism with which the protein is normally associated in nature. It is apparent that 5 the term denotes proteins of various levels of purity. Typically, a composition containing a purified protein will be one in which at least about 35%, preferably at least about 40-50%, more preferably, at least about 75-85%, and most preferably at least about 90% or more, of the total protein in the composition will be the protein in question.

By "Env polypeptide" is meant a molecule derived from an envelope protein, 10 preferably from HIV Env. The envelope protein of HIV-1 is a glycoprotein of about 160 kd (gp160). During virus infection of the host cell, gp160 is cleaved by host cell proteases to form gp120 and the integral membrane protein, gp41. The gp41 portion is anchored in (and spans) the membrane bilayer of virion, while the gp120 segment protrudes into the surrounding environment. As there is no covalent attachment between gp120 and gp41, free 15 gp120 is released from the surface of virions and infected cells. Env polypeptides may also include gp140 polypeptides. Env polypeptides can exist as monomers, dimers or multimers.

By a "gp120 polypeptide" is meant a molecule derived from a gp120 region of the Env polypeptide. Preferably, the gp120 polypeptide is derived from HIV Env. The primary 20 amino acid sequence of gp120 is approximately 511 amino acids, with a polypeptide core of about 60,000 daltons. The polypeptide is extensively modified by N-linked glycosylation to increase the apparent molecular weight of the molecule to 120,000 daltons. The amino acid sequence of gp120 contains five relatively conserved domains interspersed with five hypervariable domains. The positions of the 18 cysteine residues in the gp120 primary 25 sequence of the HIV-1_{HXB-2} (hereinafter "HXB-2") strain, and the positions of 13 of the approximately 24 N-linked glycosylation sites in the gp120 sequence are common to most, if not all, gp120 sequences. The hypervariable domains contain extensive amino acid substitutions, insertions and deletions. Despite this variation, most, if not all, gp120 sequences preserve the virus's ability to bind to the viral receptor CD4. A "gp120 polypeptide" includes both single subunits or multimers.

Env polypeptides (e.g., gp120, gp140 and gp160) include a "bridging sheet" 30 comprised of 4 anti-parallel β -strands (β -2, β -3, β -20 and β -21) that form a β -sheet. Extruding from one pair of the β -strands (β -2 and β -3) are two loops, V1 and V2. The β -2

sheet occurs at approximately amino acid residue 119 (Cys) to amino acid residue 123 (Thr) while β -3 occurs at approximately amino acid residue 199 (Ser) to amino acid residue 201 (Ile), relative to HXB-2. The "V1/V2 region" occurs at approximately amino acid positions 126 (Cys) to residue 196 (Cys), relative to HXB-2. (see, e.g., Wyatt et al. (1995) *J. Virol.* 69:5723-5733; Stamatatos et al. (1998) *J. Virol.* 72:7840-7845). Extruding from the second pair of β -strands (β -20 and β -21) is a "small-loop" structure, also referred to herein as "the bridging sheet small loop." In HXB-2, β -20 extends from about amino acid residue 422 (Gln) to amino acid residue 426 (Met) while β -21 extends from about amino acid residue 430 (Val) to amino acid residue 435 (Tyr). In variant SF162, the Met-426 is an Arg (R) residue.

5 The "small loop" extends from about amino acid residue 427 (Trp) through 429 (Lys), relative to HXB-2. A representative diagram of gp120 showing the bridging sheet, the small loop, and V1/V2 is shown in Figure 1. In addition, alignment of the amino acid sequences of Env polypeptide gp160 of selected variants is shown, relative to HXB-2, in Figures 2A-C.

10 Furthermore, an "Env polypeptide" or "gp120 polypeptide" as defined herein is not limited to a polypeptide having the exact sequence described herein. Indeed, the HIV genome is in a state of constant flux and contains several variable domains which exhibit relatively high degrees of variability between isolates. It is readily apparent that the terms encompass Env (e.g., gp120) polypeptides from any of the identified HIV isolates, as well as newly identified isolates, and subtypes of these isolates. Descriptions of structural features

15 are given herein with reference to HXB-2. One of ordinary skill in the art in view of the teachings of the present disclosure and the art can determine corresponding regions in other HIV variants (e.g., isolates HIV_{MM}, HIV_{SF2}, HIV-1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LA1}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}), and simian

20 immunodeficiency virus (SIV). (See, e.g., *Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA; for a description of these and other related viruses), using for example, sequence comparison programs (e.g., BLAST and others described herein) or identification and

25 alignment of structural features (e.g., a program such as the "ALB" program described herein that can identify β -sheet regions). The actual amino acid sequences of the modified Env polypeptides can be based on any HIV variant.

5 Additionally, the term "Env polypeptide" (e.g., "gp120 polypeptide") encompasses proteins which include additional modifications to the native sequence, such as additional internal deletions, additions and substitutions. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through naturally occurring mutational events. Thus, for example, if the Env polypeptide is to be used in vaccine compositions, the modifications must be such that immunological activity (i.e., the ability to elicit an antibody response to the polypeptide) is not lost. Similarly, if the polypeptides are to be used for diagnostic purposes, such capability must be retained.

10 Thus, a "modified Env polypeptide" is an Env polypeptide (e.g., gp120 as defined above), which has been manipulated to delete or replace all or a part of the bridging sheet portion and, optionally, the variable regions V1 and V2. Generally, modified Env (e.g., gp120) polypeptides have enough of the bridging sheet removed to expose the CD4 binding site, but leave enough of the structure to allow correct folding (e.g., correct geometry). Thus, modifications to the β -20 and β -21 regions (between about amino acid residues 420 and 435 15 relative to HBX-2) are preferred. Additionally, modifications to the β -2 and β -3 regions (between about amino acid residues 119 (Cys) and 201 (Ile)) and modifications (e.g., truncations) to the V1 and V2 loop regions may also be made. Although not all possible β -sheet and V1/V2 modifications have been exemplified herein, it is to be understood that other disrupting modifications are also encompassed by the present invention.

20 Normally, such a modified polypeptide is capable of secretion into growth medium in which an organism expressing the protein is cultured. However, for purposes of the present invention, such polypeptides may also be recovered intracellularly. Secretion into growth media is readily determined using a number of detection techniques, including, e.g., polyacrylamide gel electrophoresis and the like, and immunological techniques such as 25 Western blotting and immunoprecipitation assays as described in, e.g., International Publication No. WO 96/04301, published February 15, 1996.

30 A gp120 or other Env polypeptide is produced "intracellularly" when it is found within the cell, either associated with components of the cell, such as in association with the endoplasmic reticulum (ER) or the Golgi Apparatus, or when it is present in the soluble cellular fraction. The gp120 and other Env polypeptides of the present invention may also be secreted into growth medium so long as sufficient amounts of the polypeptides remain

present within the cell such that they can be purified from cell lysates using techniques described herein.

5 An "immunogenic" gp120 or other Env protein is a molecule that includes at least one epitope such that the molecule is capable of either eliciting an immunological reaction in an individual to which the protein is administered or, in the diagnostic context, is capable of reacting with antibodies directed against the HIV in question.

10 By "epitope" is meant a site on an antigen to which specific B cells and/or T cells respond, rendering the molecule including such an epitope capable of eliciting an immunological reaction or capable of reacting with HIV antibodies present in a biological sample. The term is also used interchangeably with "antigenic determinant" or "antigenic determinant site." An epitope can comprise 3 or more amino acids in a spatial conformation unique to the epitope. Generally, an epitope consists of at least 5 such amino acids and, more usually, consists of at least 8-10 such amino acids. Methods of determining spatial conformation of amino acids are known in the art and include, for example, x-ray 15 crystallography and 2-dimensional nuclear magnetic resonance. Furthermore, the identification of epitopes in a given protein is readily accomplished using techniques well known in the art, such as by the use of hydrophobicity studies and by site-directed serology. See, also, Geysen et al., *Proc. Natl. Acad. Sci. USA* (1984) 81:3998-4002 (general method of rapidly synthesizing peptides to determine the location of immunogenic epitopes in a given 20 antigen); U.S. Patent No. 4,708,871 (procedures for identifying and chemically synthesizing epitopes of antigens); and Geysen et al., *Molecular Immunology* (1986) 23:709-715 (technique for identifying peptides with high affinity for a given antibody). Antibodies that recognize the same epitope can be identified in a simple immunoassay showing the ability of one antibody to block the binding of another antibody to a target antigen.

25 An "immunological response" or "immune response" as used herein is the development in the subject of a humoral and/or a cellular immune response to the Env (e.g., gp120) polypeptide when the polypeptide is present in a vaccine composition. These antibodies may also neutralize infectivity, and/or mediate antibody-complement or antibody dependent cell cytotoxicity to provide protection to an immunized host. Immunological 30 reactivity may be determined in standard immunoassays, such as a competition assays, well known in the art.

Techniques for determining amino acid sequence "similarity" are well known in the art. In general, "similarity" means the exact amino acid to amino acid comparison of two or more polypeptides at the appropriate place, where amino acids are identical or possess similar chemical and/or physical properties such as charge or hydrophobicity. A so-termed "percent similarity" then can be determined between the compared polypeptide sequences.

5 Techniques for determining nucleic acid and amino acid sequence identity also are well known in the art and include determining the nucleotide sequence of the mRNA for that gene (usually via a cDNA intermediate) and determining the amino acid sequence encoded thereby, and comparing this to a second amino acid sequence. In general, "identity" refers to

10 an exact nucleotide to nucleotide or amino acid to amino acid correspondence of two polynucleotides or polypeptide sequences, respectively.

Two or more polynucleotide sequences can be compared by determining their "percent identity." Two or more amino acid sequences likewise can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or peptide sequences, is generally described as the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981). This algorithm can be extended to use with peptide sequences using the scoring matrix developed by Dayhoff, *Atlas of Protein Sequences and Structure*, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). An implementation of this algorithm for nucleic acid and peptide sequences is provided by the Genetics Computer Group (Madison, WI) in their BestFit utility application. The default parameters for this 15 method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). Other equally suitable programs for calculating the percent identity or similarity between sequences are generally known in the art.

For example, percent identity of a particular nucleotide sequence to a reference 20 sequence can be determined using the homology algorithm of Smith and Waterman with a default scoring table and a gap penalty of six nucleotide positions. Another method of establishing percent identity in the context of the present invention is to use the MPSRCH

package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages, the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension 5 penalty of one, and a gap of six). From the data generated, the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, such as the alignment program BLAST, which can also be used with default parameters. For example, BLASTN and BLASTP can be used with the following default parameters: genetic code = standard; filter = none; strand = 10 both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spudate + PIR. Details of these programs can be found at the following internet address: <http://www.ncbi.nlm.gov/cgi-bin/BLAST>.

One of skill in the art can readily determine the proper search parameters to use for a 15 given sequence in the above programs. For example, the search parameters may vary based on the size of the sequence in question. Thus, for example, a representative embodiment of the present invention would include an isolated polynucleotide having X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least about 50% identity to Y contiguous nucleotides derived from any of the sequences described herein, (ii) X equals Y, 20 and (iii) X is greater than or equal to 6 nucleotides and up to 5000 nucleotides, preferably greater than or equal to 8 nucleotides and up to 5000 nucleotides, more preferably 10-12 nucleotides and up to 5000 nucleotides, and even more preferably 15-20 nucleotides, up to the number of nucleotides present in the full-length sequences described herein (e.g., see the Sequence Listing and claims), including all integer values falling within the above-described 25 ranges.

The synthetic expression cassettes (and purified polynucleotides) of the present invention include related polynucleotide sequences having about 80% to 100%, greater than 30 80-85%, preferably greater than 90-92%, more preferably greater than 95%, and most preferably greater than 98% sequence (including all integer values falling within these described ranges) identity to the synthetic expression cassette sequences disclosed herein (for example, to the claimed sequences or other sequences of the present invention) when the sequences of the present invention are used as the query sequence.

Computer programs are also available to determine the likelihood of certain polypeptides to form structures such as β -sheets. One such program, described herein, is the "ALB" program for protein and polypeptide secondary structure calculation and prediction. In addition, secondary protein structure can be predicted from the primary amino acid

5 sequence, for example using protein crystal structure and aligning the protein sequence related to the crystal structure (e.g., using Molecular Operating Environment (MOE) programs available from the Chemical Computing Group Inc., Montreal, P.Q., Canada). Other methods of predicting secondary structures are described, for example, in Garnier et al. (1996) *Methods Enzymol.* 266:540-553; Geourjon et al. (1995) *Comput. Applic. Biosci.* 10:681-684; Levin (1997) *Protein Eng.* 10:771-776; and Rost et al. (1993) *J. Molec. Biol.* 232:584-599.

Homology can also be determined by hybridization of polynucleotides under conditions which form stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments.

15 Two DNA, or two polypeptide sequences are "substantially homologous" to each other when the sequences exhibit at least about 80%-85%, preferably at least about 90%, and most preferably at least about 95%-98% sequence identity over a defined length of the molecules, as determined using the methods above. As used herein, substantially homologous also refers to sequences showing complete identity to the specified DNA or polypeptide sequence. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al., *supra*; *DNA Cloning, supra*; *Nucleic Acid Hybridization, supra*.

20 A "coding sequence" or a sequence which "encodes" a selected protein, is a nucleic acid sequence which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to cDNA from viral nucleotide sequences as well as 25 synthetic and semisynthetic DNA sequences and sequences including base analogs. A transcription termination sequence may be located 3' to the coding sequence.

"Control elements" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the transcription and translation of a coding sequence in a host cell. Not all of these control elements need always be present so long as the desired gene is capable of being transcribed and translated.

5 A control element "directs the transcription" of a coding sequence in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

"Operably linked" refers to an arrangement of elements wherein the components so 10 described are configured so as to perform their usual function. Thus, control elements operably linked to a coding sequence are capable of effecting the expression of the coding sequence when RNA polymerase is present. The control elements need not be contiguous with the coding sequence, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between, e.g., a 15 promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a 20 polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such 25 terms denoting prokaryotic microorganisms or eucaryotic cell lines cultured as unicellular entities, are used interchangeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the 30 parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

By "vertebrate subject" is meant any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including 5 rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered.

As used herein, a "biological sample" refers to a sample of tissue or fluid isolated 10 from an individual, including but not limited to, for example, blood, plasma, serum, fecal matter, urine, bone marrow, bile, spinal fluid, lymph fluid, samples of the skin, external secretions of the skin, respiratory, intestinal, and genitourinary tracts, samples derived from the gastric epithelium and gastric mucosa, tears, saliva, milk, blood cells, organs, biopsies and also samples of *in vitro* cell culture constituents including but not limited to conditioned 15 media resulting from the growth of cells and tissues in culture medium, e.g., recombinant cells, and cell components.

The terms "label" and "detectable label" refer to a molecule capable of detection, including, but not limited to, radioactive isotopes, fluorescers, chemiluminescers, enzymes, 20 enzyme substrates, enzyme cofactors, enzyme inhibitors, chromophores, dyes, metal ions, metal sols, ligands (e.g., biotin or haptens) and the like. The term "fluorescer" refers to a substance or a portion thereof which is capable of exhibiting fluorescence in the detectable range. Particular examples of labels which may be used with the invention include, but are not limited to fluorescein, rhodamine, dansyl, umbelliferone, Texas red, luminol, acridium 25 esters, NADPH, α - β -galactosidase, horseradish peroxidase, glucose oxidase, alkaline phosphatase and urease.

Overview

The present invention concerns modified Env polypeptide molecules (e.g., 30 glycoprotein ("gp") 120). Without being bound by a particular theory, it appears that it has been difficult to generate immunological responses against Env because the CD4 binding site is buried between the outer domain, the inner domain and the V1/V2 domains. Thus, although deletion of the V1/V2 domain may render the virus more susceptible to

neutralization by monoclonal antibody directed to the CD4 site, the bridging sheet covering most of the CD4 binding domain may prevent an antibody response. Thus, the present invention provides Env polypeptides that maintain their general overall structure yet expose the CD4 binding domain. This allows the generation of an immune response (e.g., an antibody response) to epitopes in or near the CD4 binding site.

5 Various forms of the different embodiments of the invention, described herein, may be combined.

β-Sheet Conformations

10 In the present invention, location of the β-sheet structures were identified relative to 3-D (crystal) structure of an HXB-2 crystallized Env protein (see, Example 1A). Based on this structure, constructs encoding polypeptides having replacements and or excisions which maintain overall geometry while exposing the CD4 binding site were designed. In particular, the crystal structure of HXB-2 was downloaded from the Brookhaven Database. Using the 15 default parameters of the Loop Search feature of the Biopolymer module of the Sybyl molecular modeling package, homology and fit of amino acids which could replace the native loops between β-strands yet maintain overall tertiary structure were determined. Constructs encoding the modified Env polypeptides were then designed (Example 1.B.).

Thus, the modified Env polypeptides typically have enough of the bridging sheet 20 removed to expose the CD4 groove, but have enough of the structure to allow correct folding of the Env glycoprotein. Exemplary constructs are described below.

Polypeptide Production

The polypeptides of the present invention can be produced in any number of ways 25 which are well known in the art.

In one embodiment, the polypeptides are generated using recombinant techniques, well known in the art. In this regard, oligonucleotide probes can be devised based on the known sequences of the Env (e.g., gp120) polypeptide genome and used to probe genomic or cDNA libraries for Env genes. The gene can then be further isolated using standard 30 techniques and, e.g., restriction enzymes employed to truncate the gene at desired portions of the full-length sequence. Similarly, the Env gene(s) can be isolated directly from cells and tissues containing the same, using known techniques, such as phenol extraction and the

sequence further manipulated to produce the desired truncations. *See, e.g.*, Sambrook et al., *supra*, for a description of techniques used to obtain and isolate DNA.

The genes encoding the modified (*e.g.*, truncated and/or substituted) polypeptides can be produced synthetically, based on the known sequences. The nucleotide sequence can be
5 designed with the appropriate codons for the particular amino acid sequence desired. The complete sequence is generally assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. *See, e.g.*, Edge (1981) *Nature* 292:756; Nambair et al. (1984) *Science* 233:1299; Jay et al. (1984) *J. Biol. Chem.* 259:6311; Stemmer et al. (1995) *Gene* 164:49-53.

10 Recombinant techniques are readily used to clone a gene encoding an Env polypeptide gene which can then be mutagenized *in vitro* by the replacement of the appropriate base pair(s) to result in the codon for the desired amino acid. Such a change can include as little as one base pair, effecting a change in a single amino acid, or can encompass several base pair changes. Alternatively, the mutations can be effected using a mismatched
15 primer which hybridizes to the parent nucleotide sequence (generally cDNA corresponding to the RNA sequence), at a temperature below the melting temperature of the mismatched duplex. The primer can be made specific by keeping primer length and base composition within relatively narrow limits and by keeping the mutant base centrally located. *See, e.g.*, Innis et al., (1990) *PCR Applications: Protocols for Functional Genomics*; Zoller and Smith,
20 *Methods Enzymol.* (1983) 100:468. Primer extension is effected using DNA polymerase, the product cloned and clones containing the mutated DNA, derived by segregation of the primer extended strand, selected. Selection can be accomplished using the mutant primer as a hybridization probe. The technique is also applicable for generating multiple point mutations. *See, e.g.*, Dalbie-McFarland et al. *Proc. Natl. Acad. Sci USA* (1982) 79:6409.

25 Once coding sequences for the desired proteins have been isolated or synthesized, they can be cloned into any suitable vector or replicon for expression. As will be apparent from the teachings herein, a wide variety of vectors encoding modified polypeptides can be generated by creating expression constructs which operably link, in various combinations, polynucleotides encoding Env polypeptides having deletions or mutation therein. Thus,
30 polynucleotides encoding a particular deleted VI/V2 region can be operably linked with polynucleotides encoding polypeptides having deletions or replacements in the small loop

region and the construct introduced into a host cell for polypeptide expression. Non-limiting examples of such combinations are discussed in the Examples.

Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. Examples of recombinant DNA vectors 5 for cloning and host cells which they can transform include the bacteriophage λ (*E. coli*), pBR322 (*E. coli*), pACYC177 (*E. coli*), pKT230 (gram-negative bacteria), pGV1106 (gram-negative bacteria), pLAFR1 (gram-negative bacteria), pME290 (non-*E. coli* gram-negative bacteria), pHV14 (*E. coli* and *Bacillus subtilis*), pBD9 (*Bacillus*), pJJ61 (*Streptomyces*), pUC6 (*Streptomyces*), YIp5 (*Saccharomyces*), YCp19 (*Saccharomyces*) and 10 bovine papilloma virus (mammalian cells). *See, generally, DNA Cloning: Vols. I & II, supra; Sambrook et al., supra; B. Perbal, supra.*

Insect cell expression systems, such as baculovirus systems, can also be used and are known to those of skill in the art and described in, e.g., Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987). Materials and methods for 15 baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit).

Plant expression systems can also be used to produce the modified Env proteins. Generally, such systems use virus-based vectors to transfet plant cells with heterologous 20 genes. For a description of such systems see, e.g., Porta et al., *Mol. Biotech.* (1996) 5:209-221; and Hackland et al., *Arch. Virol.* (1994) 139:1-22.

Viral systems, such as a vaccinia based infection/transfection system, as described in Tomei et al., *J. Virol.* (1993) 67:4017-4026 and Selby et al., *J. Gen. Virol.* (1993) 74:1103-1113, will also find use with the present invention. In this system, cells are first 25 transfected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the DNA of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then 30 translated into protein by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation product(s).

The gene can be placed under the control of a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator (collectively referred to herein as "control" elements), so that the DNA sequence encoding the desired Env polypeptide is transcribed into RNA in the host cell transformed by a vector containing this expression construction. The 5 coding sequence may or may not contain a signal peptide or leader sequence. With the present invention, both the naturally occurring signal peptides or heterologous sequences can be used. Leader sequences can be removed by the host in post-translational processing. *See, e.g.,* U.S. Patent Nos. 4,431,739; 4,425,437; 4,338,397. Such sequences include, but are not limited to, the TPA leader, as well as the honey bee mellitin signal sequence.

10 Other regulatory sequences may also be desirable which allow for regulation of expression of the protein sequences relative to the growth of the host cell. Such regulatory sequences are known to those of skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Other types of regulatory elements may 15 also be present in the vector, for example, enhancer sequences.

The control sequences and other regulatory sequences may be ligated to the coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector which already contains the control sequences and an appropriate restriction site.

20 In some cases it may be necessary to modify the coding sequence so that it may be attached to the control sequences with the appropriate orientation; *i.e.*, to maintain the proper reading frame. Mutants or analogs may be prepared by the deletion of a portion of the sequence encoding the protein, by insertion of a sequence, and/or by substitution of one or more nucleotides within the sequence. Techniques for modifying nucleotide sequences, such 25 as site-directed mutagenesis, are well known to those skilled in the art. *See, e.g.,* Sambrook *et al., supra; DNA Cloning*, Vols. I and II, *supra; Nucleic Acid Hybridization, supra*.

The expression vector is then used to transform an appropriate host cell. A number of mammalian cell lines are known in the art and include immortalized cell lines available from the American Type Culture Collection (ATCC), such as, but not limited to, Chinese hamster 30 ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*e.g.*, Hep G2), Vero293 cells, as well as others. Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus spp.*, will find

use with the present expression constructs. Yeast hosts useful in the present invention include *inter alia*, *Saccharomyces cerevisiae*, *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillermondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect cells for use 5 with baculovirus expression vectors include, *inter alia*, *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*.

Depending on the expression system and host selected, the proteins of the present 10 invention are produced by growing host cells transformed by an expression vector described above under conditions whereby the protein of interest is expressed. The selection of the appropriate growth conditions is within the skill of the art.

In one embodiment, the transformed cells secrete the polypeptide product into the surrounding media. Certain regulatory sequences can be included in the vector to enhance 15 secretion of the protein product, for example using a tissue plasminogen activator (TPA) leader sequence, a γ -interferon signal sequence or other signal peptide sequences from known secretory proteins. The secreted polypeptide product can then be isolated by various techniques described herein, for example, using standard purification techniques such as but not limited to, hydroxyapatite resins, column chromatography, ion-exchange chromatography, size-exclusion chromatography, electrophoresis, HPLC, immunoabsorbent 20 techniques, affinity chromatography, immunoprecipitation, and the like..

Alternatively, the transformed cells are disrupted, using chemical, physical or mechanical means, which lyse the cells yet keep the Env polypeptides substantially intact. Intracellular proteins can also be obtained by removing components from the cell wall or 25 membrane, e.g., by the use of detergents or organic solvents, such that leakage of the Env polypeptides occurs. Such methods are known to those of skill in the art and are described in, e.g., *Protein Purification Applications: A Practical Approach*, (E.L.V. Harris and S. Angal, Eds., 1990)

For example, methods of disrupting cells for use with the present invention include but are not limited to: sonication or ultrasonication; agitation; liquid or solid extrusion; heat 30 treatment; freeze-thaw; desiccation; explosive decompression; osmotic shock; treatment with lytic enzymes including proteases such as trypsin, neuraminidase and lysozyme; alkali treatment; and the use of detergents and solvents such as bile salts, sodium dodecylsulphate,

Triton, NP40 and CHAPS. The particular technique used to disrupt the cells is largely a matter of choice and will depend on the cell type in which the polypeptide is expressed, culture conditions and any pre-treatment used.

Following disruption of the cells, cellular debris is removed, generally by centrifugation, and the intracellularly produced Env polypeptides are further purified, using standard purification techniques such as but not limited to, column chromatography, ion-exchange chromatography, size-exclusion chromatography, electrophoresis, HPLC, immunoabsorbent techniques, affinity chromatography, immunoprecipitation, and the like.

For example, one method for obtaining the intracellular Env polypeptides of the present invention involves affinity purification, such as by immunoaffinity chromatography using anti-Env specific antibodies, or by lectin affinity chromatography. Particularly preferred lectin resins are those that recognize mannose moieties such as but not limited to resins derived from *Galanthus nivalis* agglutinin (GNA), *Lens culinaris* agglutinin (LCA or lentil lectin), *Pisum sativum* agglutinin (PSA or pea lectin), *Narcissus pseudonarcissus* agglutinin (NPA) and *Allium ursinum* agglutinin (AUA). The choice of a suitable affinity resin is within the skill in the art. After affinity purification, the Env polypeptides can be further purified using conventional techniques well known in the art, such as by any of the techniques described above.

It may be desirable to produce Env (e.g., gp120) complexes, either with itself or other proteins. Such complexes are readily produced by e.g., co-transfected host cells with constructs encoding for the Env (e.g., gp120) and/or other polypeptides of the desired complex. Co-transfection can be accomplished either in *trans* or *cis*, i.e., by using separate vectors or by using a single vector which bears both of the Env and other gene. If done using a single vector, both genes can be driven by a single set of control elements or, alternatively, the genes can be present on the vector in individual expression cassettes, driven by individual control elements. Following expression, the proteins will spontaneously associate. Alternatively, the complexes can be formed by mixing the individual proteins together which have been produced separately, either in purified or semi-purified form, or even by mixing culture media in which host cells expressing the proteins, have been cultured. See, International Publication No. WO 96/04301, published February 15, 1996, for a description of such complexes.

Relatively small polypeptides, i.e., up to about 50 amino acids in length, can be conveniently synthesized chemically, for example by any of several techniques that are known to those skilled in the peptide art. In general, these methods employ the sequential addition of one or more amino acids to a growing peptide chain. Normally, either the amino 5 or carboxyl group of the first amino acid is protected by a suitable protecting group. The protected or derivatized amino acid can then be either attached to an inert solid support or utilized in solution by adding the next amino acid in the sequence having the complementary (amino or carboxyl) group suitably protected, under conditions that allow for the formation of an amide linkage. The protecting group is then removed from the newly added amino acid 10 residue and the next amino acid (suitably protected) is then added, and so forth. After the desired amino acids have been linked in the proper sequence, any remaining protecting groups (and any solid support, if solid phase synthesis techniques are used) are removed sequentially or concurrently, to render the final polypeptide. By simple modification of this general procedure, it is possible to add more than one amino acid at a time to a growing 15 chain, for example, by coupling (under conditions which do not racemize chiral centers) a protected tripeptide with a properly protected dipeptide to form, after deprotection, a pentapeptide. See, e.g., J. M. Stewart and J. D. Young, Solid Phase Peptide Synthesis (Pierce Chemical Co., Rockford, IL 1984) and G. Barany and R. B. Merrifield, The Peptides: Analysis, Synthesis, Biology, editors E. Gross and J. Meienhofer, Vol. 2, (Academic Press, 20 New York, 1980), pp. 3-254, for solid phase peptide synthesis techniques; and M. Bodansky, Principles of Peptide Synthesis, (Springer-Verlag, Berlin 1984) and E. Gross and J. Meienhofer, Eds., The Peptides: Analysis, Synthesis, Biology, Vol. 1, for classical solution 25 synthesis.

Typical protecting groups include t-butyloxycarbonyl (Boc), 9-fluorenylmethoxycarbonyl (Fmoc) benzyloxycarbonyl (Cbz); p-toluenesulfonyl (Tx); 2,4-dinitrophenyl; benzyl (Bzl); biphenylisopropylloxycarboxy-carbonyl, t-amylloxycarbonyl, isobornyloxycarbonyl, o-bromobenzyloxycarbonyl, cyclohexyl, isopropyl, acetyl, o-nitrophenylsulfonyl and the like.

Typical solid supports are cross-linked polymeric supports. These can include 30 divinylbenzene cross-linked-styrene-based polymers, for example, divinylbenzene-hydroxymethylstyrene copolymers, divinylbenzene-chloromethylstyrene copolymers and divinylbenzene-benzhydrylaminopolystyrene copolymers.

The polypeptide analogs of the present invention can also be chemically prepared by other methods such as by the method of simultaneous multiple peptide synthesis. See, e.g., Houghten *Proc. Natl. Acad. Sci. USA* (1985) 82:5131-5135; U.S. Patent No. 4,631,211.

5 Diagnostic and Vaccine Applications

The intracellularly produced Env polypeptides of the present invention, complexes thereof, or the polynucleotides coding therefor, can be used for a number of diagnostic and therapeutic purposes. For example, the proteins and polynucleotides or antibodies generated against the same, can be used in a variety of assays, to determine the presence of reactive 10 antibodies/and or Env proteins in a biological sample to aid in the diagnosis of HIV infection or disease status or as measure of response to immunization.

The presence of antibodies reactive with the Env (e.g., gp120) polypeptides and, conversely, antigens reactive with antibodies generated thereto, can be detected using 15 standard electrophoretic and immunodiagnostic techniques, including immunoassays such as competition, direct reaction, or sandwich type assays. Such assays include, but are not limited to, western blots; agglutination tests; enzyme-labeled and mediated immunoassays, such as ELISAs; biotin/avidin type assays; radioimmunoassays; immunoelectrophoresis; immunoprecipitation, etc. The reactions generally include revealing labels such as 20 fluorescent, chemiluminescent, radioactive, or enzymatic labels or dye molecules, or other methods for detecting the formation of a complex between the antigen and the antibody or antibodies reacted therewith.

Solid supports can be used in the assays such as nitrocellulose, in membrane or microtiter well form; polyvinylchloride, in sheets or microtiter wells; polystyrene latex, in beads or microtiter plates; polyvinylidene fluoride; diazotized paper; nylon membranes; 25 activated beads, and the like.

Typically, the solid support is first reacted with the biological sample (or the gp120 proteins), washed and then the antibodies, (or a sample suspected of containing antibodies), applied. After washing to remove any non-bound ligand, a secondary binder moiety is added 30 under suitable binding conditions, such that the secondary binder is capable of associating selectively with the bound ligand. The presence of the secondary binder can then be detected using techniques well known in the art. Typically, the secondary binder will comprise an antibody directed against the antibody ligands. A number of anti-human immunoglobulin

(Ig) molecules are known in the art (e.g., commercially available goat anti-human Ig or rabbit anti-human Ig). Ig molecules for use herein will preferably be of the IgG or IgA type, however, IgM may also be appropriate in some instances. The Ig molecules can be readily conjugated to a detectable enzyme label, such as horseradish peroxidase, glucose oxidase, 5 Beta-galactosidase, alkaline phosphatase and urease, among others, using methods known to those of skill in the art. An appropriate enzyme substrate is then used to generate a detectable signal.

Alternatively, a "two antibody sandwich" assay can be used to detect the proteins of the present invention. In this technique, the solid support is reacted first with one or more of 10 the antibodies directed against Env (e.g., gp120), washed and then exposed to the test sample. Antibodies are again added and the reaction visualized using either a direct color reaction or using a labeled second antibody, such as an anti-immunoglobulin labeled with horseradish peroxidase, alkaline phosphatase or urease.

Assays can also be conducted in solution, such that the viral proteins and antibodies 15 thereto form complexes under precipitating conditions. The precipitated complexes can then be separated from the test sample, for example, by centrifugation. The reaction mixture can be analyzed to determine the presence or absence of antibody-antigen complexes using any of a number of standard methods, such as those immunodiagnostic methods described above.

The modified Env proteins, produced as described above, or antibodies to the 20 proteins, can be provided in kits, with suitable instructions and other necessary reagents, in order to conduct immunoassays as described above. The kit can also contain, depending on the particular immunoassay used, suitable labels and other packaged reagents and materials (i.e. wash buffers and the like). Standard immunoassays, such as those described above, can be conducted using these kits.

25 The Env polypeptides and polynucleotides encoding the polypeptides can also be used in vaccine compositions, individually or in combination, in e.g., prophylactic (i.e., to prevent infection) or therapeutic (to treat HIV following infection) vaccines. The vaccines can comprise mixtures of one or more of the modified Env proteins (or nucleotide sequences encoding the proteins), such as Env (e.g., gp120) proteins derived from more than one viral 30 isolate. The vaccine may also be administered in conjunction with other antigens and immunoregulatory agents, for example, immunoglobulins, cytokines, lymphokines, and chemokines, including but not limited to IL-2, modified IL-2 (cys125~ser125), GM-CSF, IL-

12, γ -interferon, IP-10, MIP1 β and RANTES. The vaccines may be administered as polypeptides or, alternatively, as naked nucleic acid vaccines (e.g., DNA), using viral vectors (e.g., retroviral vectors, adenoviral vectors, adeno-associated viral vectors) or non-viral vectors (e.g., liposomes, particles coated with nucleic acid or protein). The vaccines may also 5 comprise a mixture of protein and nucleic acid, which in turn may be delivered using the same or different vehicles. The vaccine may be given more than once (e.g., a "prime" administration followed by one or more "boosts") to achieve the desired effects. The same composition can be administered as the prime and as the one or more boosts. Alternatively, different compositions can be used for priming and boosting.

10 The vaccines will generally include one or more "pharmaceutically acceptable excipients or vehicles" such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

15 A carrier is optionally present which is a molecule that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Furthermore, the Env 20 polypeptide may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, etc.

25 Adjuvants may also be used to enhance the effectiveness of the vaccines. Such adjuvants include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (International Publication No. WO 90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y

30 microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size

emulsion, and (c) RibiTM adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphoryl lipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (DetoxTM); (3) saponin adjuvants, 5 such as StimulonTM (Cambridge Bioscience, Worcester, MA) may be used or particle generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freunds Adjuvant (CFA) and Incomplete Freunds Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; (6) detoxified mutants of a bacterial ADP-ribosylating toxin such as a 10 cholera toxin (CT), a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63 (where lysine is substituted for the wild-type amino acid at position 63) LT-R72 (where arginine is substituted for the wild-type amino acid at position 72), CT-S109 (where serine is substituted for the wild-type amino acid at position 109), and PT-K9/G129 (where lysine is substituted for the wild-type amino acid at position 9 and glycine substituted at 15 position 129) (see, e.g., International Publication Nos. W093/13202 and W092/19265); and (7) other substances that act as immunostimulating agents to enhance the effectiveness of the composition.

20 Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetyl muramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(l'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

25 Typically, the vaccine compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above.

The vaccines will comprise a therapeutically effective amount of the modified Env 30 proteins, or complexes of the proteins, or nucleotide sequences encoding the same, and any other of the above-mentioned components, as needed. By "therapeutically effective amount" is meant an amount of a modified Env (e.g., gp120) protein which will induce a protective immunological response in the uninfected, infected or unexposed individual to which it is administered. Such a response will generally result in the development in the subject of a secretory, cellular and/or antibody-mediated immune response to the vaccine. Usually, such

a response includes but is not limited to one or more of the following effects; the production of antibodies from any of the immunological classes, such as immunoglobulins A, D, E, G or M; the proliferation of B and T lymphocytes; the provision of activation, growth and differentiation signals to immunological cells; expansion of helper T cell, suppressor T cell, 5 and/or cytotoxic T cell.

Preferably, the effective amount is sufficient to bring about treatment or prevention of disease symptoms. The exact amount necessary will vary depending on the subject being treated; the age and general condition of the individual to be treated; the capacity of the individual's immune system to synthesize antibodies; the degree of protection desired; the 10 severity of the condition being treated; the particular Env polypeptide selected and its mode of administration, among other factors. An appropriate effective amount can be readily determined by one of skill in the art. A "therapeutically effective amount" will fall in a relatively broad range that can be determined through routine trials.

Once formulated, the nucleic acid vaccines may be accomplished with or without viral 15 vectors, as described above, by injection using either a conventional syringe or a gene gun, such as the Accell® gene delivery system (PowderJect Technologies, Inc., Oxford, England). Delivery of DNA into cells of the epidermis is particularly preferred as this mode of administration provides access to skin-associated lymphoid cells and provides for a transient presence of DNA in the recipient. Both nucleic acids and/or peptides can be injected either 20 subcutaneously, epidermally, intradermally, intramuscularly such as nasally, rectally and vaginally, intraperitoneally, intravenously, orally or intramuscularly. Other modes of administration include oral and pulmonary administration, suppositories, needle-less injection, transcutaneous and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. Administration of nucleic acids may also be 25 combined with administration of peptides or other substances.

While the invention has been described in conjunction with the preferred specific 30 embodiments thereof, it is to be understood that the foregoing description as well as the examples which follow are intended to illustrate and not limit the scope of the invention. Other aspects, advantages and modifications within the scope of the invention will be apparent to those skilled in the art to which the invention pertains.

Experimental

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

5 Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

EXAMPLE 1

10 A.1. Best-Fit and Homology Searches

The crystal structure of HXB-2 gp 120 was downloaded from the Brookhaven database (COMPLEX (HIV ENVELOPE PROTEIN/CD4/FAB) 15-JUN-98 1GC1 TITLE: HIV-1 GP120 CORE COMPLEXED WITH CD4 AND A NEUTRALIZING HUMAN ANTIBODY). Beta strands 3, 2, 21, and 20 of gp 120 form a sheet near the CD4 binding site. Strands β -3 and β -2 are connected by the V1/V2 loop. Strands β -21 and β -20 are connected by another small loop. The H-bonds at the interface between strands β -2 and β -21 are the only connection between domains of the "lower" half of the protein (joining helix alpha 1 to the CD4 binding site). This beta sheet and these loops mask some antigens (e.g., antigens which may generate neutralizing antibodies) that are only exposed during the 15 CD4 binding.

20 Constructs that remove enough of the beta sheet to expose the antigens in the CD4 binding site, but leave enough of the protein to allow correct folding were designed. Specifically targeted were modifications to the small loop and, optional deletion of the V1/V2 loops. Three different types of constructs were designed: (1) constructs encoding 25 polypeptides that leave the number of residues making up the entire 4-strand beta sheet intact, but replace one or more residues; (2) constructs that encode polypeptide having at least one residue of at least one beta strand excised or (3) constructs encoding polypeptides having at least two residues of at least one beta strand excised. Thus, a total of 6 different turns were needed to rejoin the ends of the strands.

30 Initially, residues in the small loop (residues 427-430, relative to HXB-2) and connected beta strands (β -20 and β -21) were modified to contain Gly and Pro (common in beta turns). These sequences were then used as the target to match in each search. The

geometry of the target was matched to known proteins in the Brookhaven Protein Data Bank. In particular, 5-residue turns (including an overlapping single residue at the N-terminal, the 2 residue target turn and 2 overlapping residues at the C-terminal) were searched in the databases. In other words, these modified loops add a 2 residue turn that should be able to support a geometry that will maintain the beta-sheet structure of the wild type protein. The calculations were performed using the default parameters in the Loop Search feature of the Biopolymer module of the Sybyl molecular modeling package. In each case, the 25 best fits based on geometry alone were reviewed and, of those, several selected for homology and fit.

In addition, it was also determined what modifications could be made to remove most of the V1/V2 loop (residues 124-198, relative to HXB-2) yet leave the geometry of the protein intact. As with the small loop, constructs were also designed which excised one or more residues from the β -2 strand (residues 119-123 of HXB-2), the β -3 strand (residues 199-201 of HXB-2) or both β -2 and β -3. For these constructs, known loops were searched to match the geometry of a pentamer (including two remaining residues from the N-terminal side, a 2 residue turn and 1 C-terminal residue). For these searches, Gly-Gly was preferred as the insert along with at least one C-terminal substitution.

A.2. Small Loop Replacements

In one aspect, the native sequence was replaced with residues that expose the CD4 binding site, but leave the overall geometry of the protein relatively unchanged. For the small loop replacements, the target to match was: ASN425-MET426-GLY427-GLY428-GLY431. Results of the search are summarized in Table 1.

Table 1: Search of Small Loop (Asn425 through Gly431)

25

Rank	Sequence	RMSD	% Homology	Seq Id No.
Best fit	LYS-ASP-SER-ASN-ASN	0.16689	62.5	27
3	TYR-GLY-LEU-GLY-LEU	0.220308	62.5	28
4	GLU-ARG-GLU-ASP-GLY	0.241754	62.5	29
7	ARG-LYS-GLY-GLY-ASN	0.24881	100	30
12	TRP-THR-GLY-SER-TYR	0.26417	83.33	31

Based on these results, constructs encoding Gly-Gly (#7), Gly-Ser (#12) or Gly-Gly-Asn (#7) were recommended.

As V1/V2 and one or more residues of β -2 and β -3 are also optionally deleted in the modified polypeptides of the invention, known loops to match the geometry of the V1/V2 loop were also searched. The V1/V2 loop the target to match was: Lys121-Leu-122-Gly123-Gly124-Ser199. Some notable matches are shown in Table 2:

Table 2: Search of V1/V2 loop (Lys121 through Ser199)

Rank	Sequence	RMSD	% Homology	Seq Id. No.
Best fit	GLN-VAL-HIS-ASP-GLU	0.154764	68.75	32
2	LYS-GLU-GLY-ASP-LYS	0.15718	81.25	33
9	ARG-SER-GLY-ARG-SER	0.173731	68.75	34
11	THR-LEU-GLY-ASN-SER	0.175554	81.25	35
16	HIS-PHE-GLY-ALA-GLY	0.178772	93.75	36

15

Based on these searches, constructs encoding Gly-Asn in place of V1/V2 were recommended.

A.3. One Additional Residue Excisions

20

For a slightly truncated small loop, one more residue was trimmed from each beta strand to slightly shorten the beta sheet. The target to match was: ILE424-ASN425-GLY426-GLY427-LYS432. Results are shown in Table 3:

Table 3: Search of Beta sheet shortened by One residue (Ile424 through Lys432)

Rank	Sequence	RMSD	% Homology	Seq Id No.
Best fit:	ARG-MET-ALA-PRO-VAL	0.316805	58.33	37
Best hom:	ASP-SER-ASP-GLY-PRO	0.440896	83.33	38

Although these searches showed more variation and worse fits than the previous truncation, the Pro-Val or Pro-Leu encoding constructs were very similar. Accordingly, Ala-Pro encoding constructs were recommended.

Sequences encoding gp120 polypeptides having V1/V2 deleted and an additional residue from β -2 or β -3 excised were also searched. The V1/V2 loop the target to match was: VAL120-LYS121-GLY122-GLY123-VAL200. Some notable matches are shown in Table 4.

Table 4: Search of V1/V2 loop (Val120 through Val200)

Rank	Sequence	RMSD	% Homology	Seq Id No
Best fit:	THR-VAL-ASP-PRO-TYR	0.400892	58.33333	39
2	SER-THR-ASN-PRO-LEU	0.402575	54.16667	40
3	THR-ARG-SER-PRO-LEU	0.403965	58.33333	41
7	ARG-MET-ALA-PRO-VAL	0.440118	58.33333	42

15

The construct encoding Ala-Pro (e.g., #7) was recommended.

A.4. Further Excisions

In yet another truncation, an additional residue was trimmed from the β -20 and β -21 strands to further shorten the beta sheet. The target to match was ILE423-ILE424-GLY425-GLY426-ALA433. Notable matches are shown in Table 5.

Table 5: Search of Beta sheet shortened by Two Residues (Ile423 through Ala433)

Rank	Sequence	RMSD	% Homology	Seq Id No
Best fit:	THR-TYR-GLU-GLY-VAL	0.130107	79.16666	43
2	GLN-VAL-GLY-ASN-THR	0.138245	79.16666	44
3:	THR-VAL-GLY-GLY-ILE	0.153362	100	45

25

A construct encoding Gly-Gly (e.g., #3), which has 100% homology, was recommended.

30

Also searched were sequences encoding a deleted V1/V2 region and at least two residues excised from β -2, β -3 or at least one residue excised from β -2 and β -3. The target to match was: CYS119-VAL120-GLY121-GLY122-ILE201. Notable matches are shown in Table 6.

5

Table 6: Search of V1/V2 loop (Cys119 through Ile201)

Rank	Sequence	RMSD	% Homology	Seq Id No
Best fit:	ASP-LEU-PRO-GLY-CYS	0.250501	75	46
4	ASP-VAL-GLY-GLY-LEU	0.290383	100	47

10

It was determined that both constructs would be used.

B.1. Constructs encoding modified Env polypeptides

As described above, the native loops extruding from the 4- β antiparallel-stands were excised and replaced with 1 to 3 residue turns. The loops were replaced so as to leave the entire β -strands or excised by trimming one or more amino acid from each side of the connected strands. The ends of the strands were rejoined with turns that preserve the same backbone geometry (e.g., tertiary structure of β -20 and β -21), as determined by searching the Brookhaven Protein Data Bank.

20 Table 7A is a summary of the truncations of the variable regions 1 and 2 recommended for this study, as determined in Example 1.A. above.

Table 7A

V1/V2 Modifications	SEQ ID NO	Figure
-LEU122-GLY-ASN-SER199	7	10
-LYS121-ALA-PRO-VAL200-	6	9
-VAL120-GLY-GLY-ILE201-	4	7
-VAL120-PRO-GLY-ILE201B-	5	8
-VAL120-GLY-ALA-GLY-ALA204-	3	6
-VAL120-GLY-GLY-ALA-THR202-	8	11
-VAL127-GLY-ALA-GLY-ASN195-	25	28

10

As previously noted, the polypeptides encoded by the constructs of the present invention are numbered relative to HXB-2, but the particular amino acid residue of the polypeptides encoded by these exemplary constructs is based on SF-162. Thus, for example, although amino acid residue 195 in HXB-2 is a serine (S), constructs encoding polypeptides having then wild type SF162 sequence will have an asparagine (N) at this position. Table 7B shows just three of the variations in amino acid sequence between strains HXB-2 and SF162. The entire sequences, including differences in residue and amino acid number, of HXB-2 and SF162 are shown in the alignment of Figure 2 (SEQ ID NOs:1 and 2).

20

Table 7B

HXB-2 amino acid number	HXB-2 Residue	SF162 Residue/amino acid number
128	Serine (S)	Thr (T)/114
195	Serine (S)	Asn (N)/188
426	Met (M)	Arg (R)/411

25

Constructs containing deletions in the β -20 strand, β -21 stand and small loop were also constructed. Shown in Table 8 are constructs encoding truncations in these regions. The constructs in Table 8 are numbered relative to HXB-2 but the unmodified amino acid sequence is based on SF162. Thus, the construct encodes an arginine (Arg) as is found in

SF162 in the amino acid position numbered 426 relative to HXB-2 (See, also, Table 7B). Changes from wildtype (SF162) are shown in bold in Table 8B.

Table 8

	Small Loop/β-20 and β-21 (Modified)	SEQ ID NO	Figure
5	-TRP427-GLY-GLY431-	9	12
	-ARG426-GLY-GLY-GLY431-	10	13
	-ARG426-GLY-SER-GLY431B-	11	14
	-ARG426-GLY-GLY-ASN-LYS432-	12	15
	-ASN425-ALA-PRO-LYS432-	13	16
10	-ILE424-GLY-GLY-ALA433-	14	17
	-ILE423-GLY-GLY-MET434-	15	18
	GLN422-GLY-GLY-TYR435-	16	19
	-GLN422-ALA-PRO-TYR435B-	17	20

15

The deletion constructs shown in Tables 7 and 8 for each one of the β-strands and combinations of them are constructed. These deletions will be tested in the Env forms gp120, gp140 and gp160 from different HIV strains like subtype B strains (e.g., SF162, US4, SF2), subtype E strains (e.g., CM235) and subtype C strains (e.g., AF110968 or AF110975).

20 Exemplary constructs for SF162 are shown in the

Figures and are summarized in Table 9. As noted above in Figure 2 and Table 7B, in the bridging sheet region, the amino acid sequence of SF162 differs from HXB-2 in that the Met426 of HXB-2 is an Arg in SF162. In Table 9, V1/V2 refers to deletions in the V1/V2 region; # bsm refers to a modification in the bridging sheet small loop.

25

Table 9

Construct	Seq. Id.	Fig.	Modification/Amino acid sequence
Val120-Ala204	3	6	V1/V2: Val120-Gly-Ala-Gly-Ala204
Val120-Ile201	4	7	V1/V2: Val120-Gly-Gly-Ile201
Val120-Ile201B	5	8	V1/V2: Val120-Pro-Gly-Ile201
Lys121-Val200	6	9	V1/V2: Lys121-Ala-Pro-Val200

Table 9

Construct	Seq. Id.	Fig.	Modification/Amino acid sequence
Leu122-Ser199	7	10	V1/V2: Leu122-Gly-Asn-Ser199
Val120-Thr202	8	11	V1/V2: Val120-Gly-Gly-Ala-Thr202
Trp427-Gly431	9	12	bsm: Trp427-Gly-Gly431
Arg426-Gly431	10	13	bsm: Arg426-Gly-Gly-Gly431
5 Arg426-Gly431B	11	14	bsm: Arg426-Gly-Ser-Gly431
Arg426-Lys432	12	15	bsm: Arg426-Gly-Gly-Asn-Lys432
Asn425-Lys432	13	16	bsm: Asn425-Ala-Pro-Lys432
Ile424-Ala433	14	17	bsm: Ile424-Gly-Gly-Ala433
Ile423-Met434	15	18	bsm: Ile423-Gly-Gly-Met434
10 Gln422-Tyr435	16	19	bsm: Gln422-Gly-Gly-Tyr435
Val127-Asn195	25	28	bsm: Val127-Gly-Ala-Gly-Asn195
Gln422-Tyr435B	17	20	bsm: Gln422-Ala-Pro-Tyr435
Leu122-Ser199; Arg426-Gly431	18	21	V1/V2/bsm: Leu122-Gly-Asn-Ser199 --- Arg426-Gly-Gly-Gly431
15 Leu122-Ser199; Arg426-Lys432	19	22	V1/V2/bsm: Leu122-Gly-Asn-Ser199 --- Arg426-Gly-Gly-Asn-Lys432
Leu122-Ser199-Trp427-Gly431	20	23	V1/V2/bsm: Leu122-Gly-Asn-Ser199 --- Trp427-Gly-Gly431
20 Lys121-Val200-Asn425-Lys432	21	24	V1/V2/bsm: Lys121-Ala-Pro-Val200 --- Asn425-Ala-Pro-Lys432
Val120-Ile201-Ile424-Ala433	22	25	V1/V2/bsm: Val120-Gly-Gly-Ile201 --- Ile424-Gly-Gly-Ala433
Val120-Ile201B-Ile424-Ala433	23	26	V1/V2/bsm: Val120-Pro-Gly-Ile201 --- Ile424-Gly-Gly-Ala433
25 Val120-Thr202; Ile424-Ala433	24	27	V1/V2/bsm: Val120-Gly-Gly-Ala-Thr202 --- Ile424-Gly-Gly-Ala433
Val127-Asn195; Arg426-Gly431	25	29	V1/V2/bsm: Val127-Gly-Ala-Gly-Asn195 --- Arg426-Gly-Gly-Gly431

30

Combinations of V1/V2 deletions and bridging sheet small loop modifications in addition to those specifically shown in Table 9 are also within the scope of the present invention. Various forms of the different embodiments of the invention, described herein, may be combined.

The first screening will be done after transient expression in COS-7, RD and/or 293 cells. The proteins that are expressed will be analyzed by immunoblot, ELISA, and for binding to mAbs directed to the CD4 binding site and other important epitopes on gp120 to determine integrity of structure. They will also be tested in a CD4 binding assay and, in 5 addition, the binding of neutralizing antibodies, for example using patient sera or mAb 448D (directed to Glu370 and Tyr384, a region of the CD4 binding groove that is not altered by the deletions).

10 The immunogenicity of these novel Env glycoproteins will be tested in rodents and primates. The structures will be administered as DNA vaccines or adjuvanted protein vaccines or in combined modalities. The goal of these vaccinations will be to archive broadly reactive neutralizing antibody responses.

Claims:

What is claimed is:

- 5 1. A polynucleotide encoding a modified HIV Env polypeptide wherein the polypeptide has at least one amino acid deleted or replaced in the region corresponding to residues 420 to 436 relative to HXB-2 (SEQ ID NO:1).
- 10 2. The polynucleotide of claim 1, wherein the region corresponding to residues 124-198 relative to HXB-2 is deleted and at least one amino acid is deleted or replaced in the regions corresponding to the residues 119 to 123 and 199 to 210 relative to HXB-2 (SEQ ID NO:1).
- 15 3. The polynucleotide of claim 1, wherein at least one amino acid in the region corresponding to residues 427 through 429 relative to HXB-2 (SEQ ID NO:1) is deleted or replaced.
- 20 4. The polynucleotide of claim 2, wherein at least one amino acid of the in the region corresponding to residues 427 through 429 relative to HXB-2 (SEQ ID NO:1) is deleted or replaced.
- 25 5. The polynucleotide of claim 1, wherein the amino acid sequence of the modified HIV Env polypeptide is based on strain SF162.
- 30 6. An immunogenic modified HIV Env polypeptide having at least one amino acid deleted or replaced in the region corresponding to residues 420 through 436, relative to HXB-2 (SEQ ID NO:1).
7. The polypeptide of claim 6, wherein one amino acid is deleted in the region corresponding to residues 420 through 436, relative to HXB-2 (SEQ ID NO:1).

8. The polypeptide of claim 6, wherein more than one amino acid is deleted in the region corresponding to residues 420 through 436, relative to HXB-2 (SEQ ID NO:1).

9. The polypeptide of claim 6, wherein at least one amino acid is replaced in the 5 region corresponding to residues 420 through 436, relative to HXB-2 (SEQ ID NO:1).

10. The polypeptide of claim 6, wherein at least one amino acid residue between about amino acid residue 427 and amino acid residue 429 relative to HXB-2 (SEQ ID NO:1) is deleted or replaced.

10

11. The polypeptide of claim 6, wherein the V1 and V2 regions of the polypeptide are truncated.

15

12. The polypeptide of claim 10, wherein the V1 and V2 regions of the polypeptide are truncated.

13. The polypeptide of claim 6, wherein the amino acid sequence of the modified HIV Env polypeptide is based on strain SF162.

20

14. A construct comprising the nucleotide sequence depicted in Figure 6 (SEQ ID NO:3).

15. A construct comprising the nucleotide sequence depicted in Figure 7 (SEQ ID NO:4).

25

16. A construct comprising the nucleotide sequence depicted in Figure 8 (SEQ ID NO:5).

30

17. A construct comprising the nucleotide sequence depicted in Figure 9 (SEQ ID NO:6).

18. A construct comprising the nucleotide sequence depicted in Figure 10 (SEQ ID NO:7).

19. A construct comprising the nucleotide sequence depicted in Figure 11 (SEQ ID 5 NO:8).

20. A construct comprising the nucleotide sequence depicted in Figure 12 (SEQ ID NO:9).

10 21. A construct comprising the nucleotide sequence depicted in Figure 13 (SEQ ID NO:10).

22. A construct comprising the nucleotide sequence depicted in Figure 14 (SEQ ID NO:11).

15 23. A construct comprising the nucleotide sequence depicted in Figure 15 (SEQ ID NO:12).

20 24. A construct comprising the nucleotide sequence depicted in Figure 16 (SEQ ID NO:13).

25 25. A construct comprising the nucleotide sequence depicted in Figure 17 (SEQ ID NO:14).

26. A construct comprising the nucleotide sequence depicted in Figure 18 (SEQ ID 25 NO:15).

27. A construct comprising the nucleotide sequence depicted in Figure 19 (SEQ ID NO:16).

30 28. A construct comprising the nucleotide sequence depicted in Figure 20 (SEQ ID NO:17).

29. A construct comprising the nucleotide sequence depicted in Figure 21 (SEQ ID NO:18).

30. A construct comprising the nucleotide sequence depicted in Figure 22 (SEQ ID 5 NO:19).

31. A construct comprising the nucleotide sequence depicted in Figure 23 (SEQ ID NO:20).

10 32. A construct comprising the nucleotide sequence depicted in Figure 24 (SEQ ID NO:21).

33. A construct comprising the nucleotide sequence depicted in Figure 25 (SEQ ID NO:22).

15 34. A construct comprising the nucleotide sequence depicted in Figure 26 (SEQ ID NO:23).

20 35. A construct comprising the nucleotide sequence depicted in Figure 27 (SEQ ID NO:24).

36. A construct comprising the nucleotide sequence depicted in Figure 28 (SEQ ID NO:25).

25 37. A construct comprising the nucleotide sequence depicted in Figure 29 (SEQ ID NO:26).

38. A vaccine composition comprising a polynucleotide encoding a modified Env polypeptide according to any one of claims 1-5.

30 39. A vaccine composition comprising a polynucleotide construct encoding a modified Env polypeptide according to any of claims 14-37.

40. A vaccine composition comprising a modified Env polypeptide according to any of claims 6-13.

5 41. The vaccine composition of any of claims 38-40, further comprising an adjuvant.

10 42. A method of inducing an immune response in subject comprising, administering a polynucleotide according to any one of claims 1-5 in an amount sufficient to induce an immune response in the subject.

15 43. A method of inducing an immune response in subject comprising, administering a polynucleotide construct according to any one of claims 14-37 in an amount sufficient to induce an immune response in the subject.

15 44. A method of inducing an immune response in a subject comprising administering a composition comprising a modified Env polypeptide according to any one of claims 6-13, wherein the composition is administered in an amount sufficient to induce an immune response in the subject

20 45. The method of any of claims 42-44 further comprising administering an adjuvant to the subject.

25 46. A method of inducing an immune response in a subject comprising
 (a) administering a first composition comprising a polynucleotide according to any of claims 1-5 in a priming step and
 (b) administering a second composition comprising a modified Env polypeptide according to any of claims 6-13, as a booster, in an amount sufficient to induce an immune response in the subject.

30 47. The method of claim 46 wherein the first composition or second composition further comprise an adjuvant.

48. The method of claim 46 wherein the first and second compositions further comprise an adjuvant.

gp120 core structure

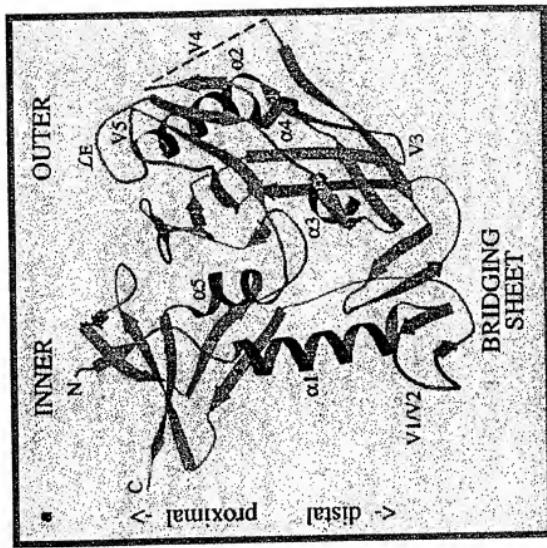


FIG. 1

1 50 100 150 200 250 300 350

HXB2 (1) MRVK---EK**Y**QHLWRW**G**WRWGT**L**LL**G**LMIC-S**T**E**K**W**T**Y**W**PK
 162 (1) ---MD**A**RC**C**CC**V**LL**L**CG**A**F**S**P**S**VE**K**E**D**Y**W**PK
 SF2 (1) MKVKGTRRN**Q**HLWRW**G**---T**L**LL**G**LMIC-S**T**E**K**W**T**Y**W**PK
 CM236 (1) M**R**V**K**E**T**Q**M**N**P**N**L****W****G**---T**L**LL**G**LMIC-S**T**E**K**W**T**Y**W**PK
 US4 (1) ---**W**---K**C**Q**H**LWRW**G**---T**L**LL**G**LMIC-S**T**E**K**W**T**Y**W**PK
 Consensus (1) MRVK YOHLWRWG TLLGLMILIC SATEKILWTVTYYGGPVWK

101 150

HXB2 (47) EAT**T**IL**F**CAS**D**AK**Y**K**Y**D**T**E**V**H**N**V**W**A**T**H**C**V**T**D**P**N**P**QE**V**V**L** N**V**T**E**N**F**N**M**
 162 (41) EAT**T**IL**F**CAS**D**AK**Y**K**Y**D**T**E**V**H**N**V**W**A**T**H**C**V**T**D**P**N**P**QE**V**V**L** N**V**T**E**N**F**N**M**
 SF2 (46) EAT**T**IL**F**CAS**D**AK**Y**K**Y**D**T**E**V**H**N**V**W**A**T**H**C**V**T**D**P**N**P**QE**V**V**L** N**V**T**E**N**F**N**M**
 CM236 (46) EAT**T**IL**F**CAS**D**AK**Y**K**Y**D**T**E**V**H**N**V**W**A**T**H**C**V**T**D**P**N**P**QE**V**V**L** N**V**T**E**N**F**N**M**
 US4 (41) EAT**T**IL**F**CAS**D**AK**Y**K**Y**D**T**E**V**H**N**V**W**A**T**H**C**V**T**D**P**N**P**QE**V**V**L** N**V**T**E**N**F**N**M**
 Consensus (51) EAT**T**IL**F**CAS**D**AK**Y**K**Y**D**T**E**V**H**N**V**W**A**T**H**C**V**T**D**P**N**P**QE**V**V**L** N**V**T**E**N**F**N**M**

101 150

HXB2 (97) KND**W**Q**P**N**E****H****I****L****S****T****N****S****G****M****I****N****E****R****G****E****I****K****R****E****N****I****T****S****I****D****K****O****L**-----
 162 (91) KNN**W**Q**P**N**E****H****I****L****S****T****N****S****G****M****I****N****E****R****G****E****I****K****R****E****N****I****T****S****I****D****K****O****L**-----
 SF2 (96) KNN**W**Q**P**N**E****H****I****L****S****T****N****S****G****M****I****N****E****R****G****E****I****K****R****E****N****I****T****S****I****D****K****O****L**-----
 CM236 (96) KNN**W**Q**P**N**E****H****I****L****S****T****N****S****G****M****I****N****E****R****G****E****I****K****R****E****N****I****T****S****I****D****K****O****L**-----
 US4 (91) KNN**W**Q**P**N**E****H****I****L****S****T****N****S****G****M****I****N****E****R****G****E****I****K****R****E****N****I****T****S****I****D****K****O****L**-----
 Consensus (101) KNN**W**Q**P**N**E****H****I****L****S****T****N****S****G****M****I****N****E****R****G****E****I****K****R****E****N****I****T****S****I****D****K****O****L****T****G****S****T****N****G****T****N****S****T****S**

151 200

HXB2 (135) -----KND**T**NT**N**SS**G****M****I****N****E****R****G****E****I****K****R****E****N****I****T****S****I****D****K****O****L**-----
 162 (129) -----K**N**T**N**T**N**SS**G****M****I****N****E****R****G****E****I****K****R****E****N****I****T****S****I****D****K****O****L**-----
 SF2 (134) -----G**K**T**N**T**N**SS**G****M****I****N****E****R****G****E****I****K****R****E****N****I****T****S****I****D****K****O****L**-----
 CM236 (135) -----L**T**N**V**N**T**N**I****V**S**N**T**I****G**N**I****T****D****E****N****S****I****T****E****L****D****I****D****K****O****V****A****L****Y**
 US4 (141) GT**N**T**S**GT**N**T**S**NT**S**DS**W**E**K****S**PE**E****I****K****R****E****N****I****T****S****I****D****K****O****V****A****L****Y**
 Consensus (151) NATNTNSS KE M G**E****I****K****N**C**S**F**I****T****S****I****R****D****K****V****Q****K****E****Y****A****L****Y**

201 250

HXB2 (178) K**L**D**V****P**D**N**D**T****S**-----**T****S****E****T****S****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 162 (171) K**L**D**V****P**D**N**D**T****S**-----**T****S****E****T****S****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 SF2 (176) N**D**V**V****P**D**N**D**T****S**-----**T****S****E****T****S****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 CM236 (179) K**L**D**V****P**D**N**D**T****S**-----**S****E****R****E****I****N****T****M****K****D****R****C****S****H****E****P****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 US4 (191) K**L**D**V****P**D**N**D**T****S**-----**T****S****E****T****S****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 Consensus (201) KLDVVPIDND TS YRLINCNTSVITQACPKVSFEPPIHYCAPG

251 300

HXB2 (223) F**A****I****L****C****N****D****K****F****G****T****N****S**-----**T****S****E****T****S****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 162 (216) F**A****I****L****C****N****D****K****F****G****T****N****S**-----**T****S****E****T****S****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 SF2 (226) F**A****I****L****C****N****D****K****F****G****T****N****S**-----**T****S****E****T****S****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 CM236 (226) F**A****I****L****C****N****D****K****F****G****T****N****S**-----**T****S****E****T****S****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 US4 (236) F**A****I****L****C****N****D****K****F****G****T****N****S**-----**T****S****E****T****S****I****N****T****S****M****I****T****T****A****C****R****V****S****E****E****I****H****G****A****P**
 Consensus (251) FAILKNDK FNGTGCTNVSTVQCTHIGRPPVSTQLLNGLAEEEVVI

301 350

HXB2 (273) RSVNFD**D****I****N****V****E****O****N****T****V****E****I****N****T****N****E****N****T****H****K****R****E****P****R****O****G****R****A****V****T****I****S****K**
 162 (266) RSENFTD**D****I****N****V****E****O****N****T****V****E****I****N****T****N****E****N****T****H****K****R****E****P****R****O****G****R****A****V****T****I****S****K**
 SF2 (276) RSDNFTD**D****I****N****V****E****O****N****T****V****E****I****N****T****N****E****N****T****H****K****R****E****P****R****O****G****R****A****V****T****I****S****K**
 CM236 (276) RSENFTD**D****I****N****V****E****O****N****T****V****E****I****N****T****N****E****N****T****H****K****R****E****P****R****O****G****R****A****V****T****I****S****K**
 US4 (286) RSENFTD**D****I****N****V****E****O****N****T****V****E****I****N****T****N****E****N****T****H****K****R****E****P****R****O****G****R****A****V****T****I****S****K**
 Consensus (301) RSENFTDANKTIVQLNESVEINCTRPNNNTRSKI I GPGRAFY TGD

FIG. 2A

FIG. 2B

		651	700
HXB2	(618)	SLEQIHNNTWMEWEREINNTSLIHSLLIES	QNEQEKDEQE
162	(604)	SLEQIHNNTWMEWEREINNTSLIHSLLIES	QNEQEKDEQE
SF2	(617)	SLEQIHNNTWMEWEREINNTSLIHSLLIES	QNEQEKDEQE
CM236	(613)	SYEHNNTWMEWEREINNTSLIHSLLIES	QNEQERNEKD
US4	(629)	SLETWDNMIDMEWEREINNTSLIHSLLIES	QNEQERNEKD
Consensus	(651)	SLEEIWNNTWMEWEREIN	QNEQEKDEQE
			750
HXB2	(668)	SQHNNTNNDWMEWEREINNTSLIHSLLIES	QNEQEKDEQE
162	(654)	SQHNNTNNDWMEWEREINNTSLIHSLLIES	QNEQEKDEQE
SF2	(667)	SQHNNTNNDWMEWEREINNTSLIHSLLIES	QNEQEKDEQE
CM236	(663)	SQHNNTNNDWMEWEREINNTSLIHSLLIES	QNEQEKDEQE
US4	(679)	SQHNNTNNDWMEWEREINNTSLIHSLLIES	QNEQEKDEQE
Consensus	(701)	SILWNFDITNWLYIKIFIMIVGGLVGLRIVFAVLISIVNRVRQGYSPLSF	
			800
HXB2	(718)	WHLPTPEGIEEGEREGEGERD	RDGSILRVNGSLHLLIDDEBSLNS
162	(704)	WRFPAPPGIEEGEREGEGERD	RDGSILRVNGSLHLLIDDEBSLNS
SF2	(717)	WRLPVPKGIEEGEREGEGERD	RDGSILRVNGSLHLLIDDEBSLNS
CM236	(713)	WRFHHQEPESEREIEEGEREGEGERD	RDGSILRVNGSLHLLIDDEBSLNS
US4	(729)	WRLPAQGIEEGEREGEGERD	RDGSILRVNGSLHLLIDDEBSLNS
Consensus	(751)	QTRLP PRGPDRPEGIEEGERDRDRSVRLV	G LALIWDDLRLSCLFS
			850
HXB2	(768)	WHRDLDLLIAVIRIVELGLR-----	RGEALKYMWNLLQYWQELKNS
162	(754)	WHRDLDLLIAVIRIVELGLR-----	RGEALKYMWNLLQYWQELKNS
SF2	(767)	WHRDLDLLIAVIRIVELGLR-----	RGEALKYMWNLLQYWQELKNS
CM236	(763)	WHRDLDLLIAVIRIVELGLR-----	RGEALKYMWNLLQYWQELKNS
US4	(779)	WHRDLDLLIAVIRIVELGLR-----	RGEALKYMWNLLQYWQELKNS
Consensus	(801)	YHRLRLDLLLLIAVIRIVELGLR-----	RGEALKYMWNLLQYWQELKNS
			900
HXB2	(811)	AVSLLNNTWAVAAEGDGRV	EVVQGACQAIRIDPNSRQGLR-----
162	(797)	AVSLFDI1WAVAAEGDGRV	EVVQGACQAIRIDPNSRQGLR-----
SF2	(810)	AVSLNNTWAVAAEGDGRV	EVVQGACQAIRIDPNSRQGLR-----
CM236	(813)	AVSLLDI1WAVAAEGDGRV	EVVQGACQAIRIDPNSRQGLR-----
US4	(822)	AVSLFNNTWAVAAEGDGRV	EVVQGACQAIRIDPNSRQGLR-----
Consensus	(851)	AVSLLNATAIAVAAEGDGRV	EVVQGACQAIRIDPNSRQGLR-----

FIG. 2C

	1	50
Leu122-Ser199	{1} GAATTCCGACCATGGATGCAATGAGAGGGCTCTGCT	
Val127-Asn195	{1} GAATTCCGACCATGGATGCAATGAGAGGGCTCTGCT	
Val120-Ile2018	{1} GAATTCCGACCATGGATGCAATGAGAGGGCTCTGCT	
Val120-Ala204	{1} GAATTCCGACCATGGATGCAATGAGAGGGCTCTGCT	
Val120-Ile201	{1} GAATTCCGACCATGGATGCAATGAGAGGGCTCTGCT	
Val120-Thr202	{1} GAATTCCGACCATGGATGCAATGAGAGGGCTCTGCT	
Lys121-Val200	{1} GAATTCCGACCATGGATGCAATGAGAGGGCTCTGCT	
Consensus	{1} GAATTCCGACCATGGATGCAATGAGAGGGCTCTGCT	
	41	80
Leu122-Ser199	{41} GTGTCGTCGTCGTCGTCGAGACAGTCTCGTTCCGCCAG	
Val127-Asn195	{41} GTGTCGTCGTCGTCGTCGAGACAGTCTCGTTCCGCCAG	
Val120-Ile2018	{41} GTGTCGTCGTCGTCGTCGAGACAGTCTCGTTCCGCCAG	
Val120-Ala204	{41} GTGTCGTCGTCGTCGTCGAGACAGTCTCGTTCCGCCAG	
Val120-Ile201	{41} GTGTCGTCGTCGTCGTCGAGACAGTCTCGTTCCGCCAG	
Val120-Thr202	{41} GTGTCGTCGTCGTCGTCGAGACAGTCTCGTTCCGCCAG	
Lys121-Val200	{41} GTGTCGTCGTCGTCGTCGAGACAGTCTCGTTCCGCCAG	
Consensus	{41} GTGTCGTCGTCGTCGTCGAGACAGTCTCGTTCCGCCAG	
	81	120
Leu122-Ser199	{81} CGCGTGGAGAAAGCTGTGGGTGACCGTGTACTACGGCGTG	
Val127-Asn195	{81} CGCGTGGAGAAAGCTGTGGGTGACCGTGTACTACGGCGTG	
Val120-Ile2018	{81} CGCGTGGAGAAAGCTGTGGGTGACCGTGTACTACGGCGTG	
Val120-Ala204	{81} CGCGTGGAGAAAGCTGTGGGTGACCGTGTACTACGGCGTG	
Val120-Ile201	{81} CGCGTGGAGAAAGCTGTGGGTGACCGTGTACTACGGCGTG	
Val120-Thr202	{81} CGCGTGGAGAAAGCTGTGGGTGACCGTGTACTACGGCGTG	
Lys121-Val200	{81} CGCGTGGAGAAAGCTGTGGGTGACCGTGTACTACGGCGTG	
Consensus	{81} CGCGTGGAGAAAGCTGTGGGTGACCGTGTACTACGGCGTG	
	121	160
Leu122-Ser199	{121} CGCGTGTGGAGAGGCCACCCACCCCTGTTCTGGCCA	
Val127-Asn195	{121} CGCGTGTGGAGAGGCCACCCACCCCTGTTCTGGCCA	
Val120-Ile2018	{121} CGCGTGTGGAGAGGCCACCCACCCCTGTTCTGGCCA	
Val120-Ala204	{121} CGCGTGTGGAGAGGCCACCCACCCCTGTTCTGGCCA	
Val120-Ile201	{121} CGCGTGTGGAGAGGCCACCCACCCCTGTTCTGGCCA	
Val120-Thr202	{121} CGCGTGTGGAGAGGCCACCCACCCCTGTTCTGGCCA	
Lys121-Val200	{121} CGCGTGTGGAGAGGCCACCCACCCCTGTTCTGGCCA	
Consensus	{121} CGCGTGTGGAGAGGCCACCCACCCCTGTTCTGGCCA	
	161	200
Leu122-Ser199	{161} GCGACGCCAAAGCTTACGACACCGAGGTGCAACAGTGTG	
Val127-Asn195	{161} GCGACGCCAAAGCTTACGACACCGAGGTGCAACAGTGTG	
Val120-Ile2018	{161} GCGACGCCAAAGCTTACGACACCGAGGTGCAACAGTGTG	
Val120-Ala204	{161} GCGACGCCAAAGCTTACGACACCGAGGTGCAACAGTGTG	
Val120-Ile201	{161} GCGACGCCAAAGCTTACGACACCGAGGTGCAACAGTGTG	
Val120-Thr202	{161} GCGACGCCAAAGCTTACGACACCGAGGTGCAACAGTGTG	
Lys121-Val200	{161} GCGACGCCAAAGCTTACGACACCGAGGTGCAACAGTGTG	
Consensus	{161} GCGACGCCAAAGCTTACGACACCGAGGTGCAACAGTGTG	
	201	240
Leu122-Ser199	{201} GGCCACCCACGCCCTACGACACCGAGGTGCAACAGTGTG	
Val127-Asn195	{201} GGCCACCCACGCCCTACGACACCGAGGTGCAACAGTGTG	
Val120-Ile2018	{201} GGCCACCCACGCCCTACGACACCGAGGTGCAACAGTGTG	
Val120-Ala204	{201} GGCCACCCACGCCCTACGACACCGAGGTGCAACAGTGTG	
Val120-Ile201	{201} GGCCACCCACGCCCTACGACACCGAGGTGCAACAGTGTG	
Val120-Thr202	{201} GGCCACCCACGCCCTACGACACCGAGGTGCAACAGTGTG	
Lys121-Val200	{201} GGCCACCCACGCCCTACGACACCGAGGTGCAACAGTGTG	
Consensus	{201} GGCCACCCACGCCCTACGACACCGAGGTGCAACAGTGTG	
	241	280
Leu122-Ser199	{241} GAGATCGTCGTCGAGACGTGACCGAGAACTTCAACATGT	
Val127-Asn195	{241} GAGATCGTCGTCGAGACGTGACCGAGAACTTCAACATGT	

Val120-Ile201B	(241) GAGATCGTGGAGAACGTGACCGAGAACTTCAACATGT
Val120-Ala204	(241) GAGATCGTGGAGAACGTGACCGAGAACTTCAACATGT
Val120-Ile201	(241) GAGATCGTGGAGAACGTGACCGAGAACTTCAACATGT
Val120-Thr202	(241) GAGATCGTGGAGAACGTGACCGAGAACTTCAACATGT
Lys121-Val200	(241) GAGATCGTGGAGAACGTGACCGAGAACTTCAACATGT
Consensus	(241) GAGATCGTGGAGAACGTGACCGAGAACTTCAACATGT 281 320
Leu122-Ser199	(281) GGAAGAACAAACATGGTGGAGCAGATGCACGGAGACATCAT
Val127-Asn195	(281) GGAAGAACAAACATGGTGGAGCAGATGCACGGAGACATCAT
Val120-Ile201B	(281) GGAAGAACAAACATGGTGGAGCAGATGCACGGAGACATCAT
Val120-Ala204	(281) GGAAGAACAAACATGGTGGAGCAGATGCACGGAGACATCAT
Val120-Ile201	(281) GGAAGAACAAACATGGTGGAGCAGATGCACGGAGACATCAT
Val120-Thr202	(281) GGAAGAACAAACATGGTGGAGCAGATGCACGGAGACATCAT
Lys121-Val200	(281) GGAAGAACAAACATGGTGGAGCAGATGCACGGAGACATCAT
Consensus	(281) GGAAGAACAAACATGGTGGAGCAGATGCACGGAGACATCAT 321 360
Leu122-Ser199	(321) CAGCTGTGGGACCAAGAGCTGAAGCCCTGCGTGAAGCTG
Val127-Asn195	(321) CAGCTGTGGGACCAAGAGCTGAAGCCCTGCGTGAAGCTG
Val120-Ile201B	(321) CAGCTGTGGGACCAAGAGCTGAAGCCCTGCGTGGG----
Val120-Ala204	(321) CAGCTGTGGGACCAAGAGCTGAAGCCCTGCGTGGG----
Val120-Ile201	(321) CAGCTGTGGGACCAAGAGCTGAAGCCCTGCGTGGG----
Val120-Thr202	(321) CAGCTGTGGGACCAAGAGCTGAAGCCCTGCGTGAAGG--
Lys121-Val200	(321) CAGCTGTGGGACCAAGAGCTGAAGCCCTGCGTGAAGG--
Consensus	(321) CAGCTGTGGGACCAAGAGCTGAAGCCCTGCGTGGG 361 400
Leu122-Ser199	(361) -----GGCAA-----CGGG
Val127-Asn195	(361) ACCCCCCCTGTGCGTGGGGCAGGGAACTGCAACACAGGG
Val120-Ile201B	(357) -----CG
Val120-Ala204	(357) -----CG
Val120-Ile201	(357) -----CG
Val120-Thr202	(357) -----CG
Lys121-Val200	(359) -----C-----CCCG
Consensus	(361) -----CG 401 440
Leu122-Ser199	(371) TGATCACCCAGGCTGCCCAAGGTGAGCTTCGAGCCAT
Val127-Asn195	(401) TGATCACCCAGGCTGCCCAAGGTGAGCTTCGAGCCAT
Val120-Ile201B	(359) GCATCACCCAGGCTGCCCAAGGTGAGCTTCGAGCCAT
Val120-Ala204	(357) ---CGCCGGCAGCTGCCCAAGGTGAGCTTCGAGCCAT
Val120-Ile201	(359) GCATCACCCAGGCTGCCCAAGGTGAGCTTCGAGCCAT
Val120-Thr202	(359) GGCACCCACAGGCTGCCCAAGGTGAGCTTCGAGCCAT
Lys121-Val200	(365) TGATCACCCAGGCTGCCCAAGGTGAGCTTCGAGCCAT
Consensus	(401) ATCACCCAGGCTGCCCAAGGTGAGCTTCGAGCCAT 441 480
Leu122-Ser199	(411) CCCCCATCCACTACTGCGCCCGGCCGCTTCGCCATCCGT
Val127-Asn195	(441) CCCCCATCCACTACTGCGCCCGGCCGCTTCGCCATCCGT
Val120-Ile201B	(399) CCCCCATCCACTACTGCGCCCGGCCGCTTCGCCATCCGT
Val120-Ala204	(393) CCCCCATCCACTACTGCGCCCGGCCGCTTCGCCATCCGT
Val120-Ile201	(399) CCCCCATCCACTACTGCGCCCGGCCGCTTCGCCATCCGT
Val120-Thr202	(399) CCCCCATCCACTACTGCGCCCGGCCGCTTCGCCATCCGT
Lys121-Val200	(405) CCCCCATCCACTACTGCGCCCGGCCGCTTCGCCATCCGT
Consensus	(441) CCCCCATCCACTACTGCGCCCGGCCGCTTCGCCATCCGT 481 520
Leu122-Ser199	(451) AAGTCAACGACAAGAAGTTCAGGGAGGGGCCCTGCA
Val127-Asn195	(481) AAGTCAACGACAAGAAGTTCAGGGAGGGGCCCTGCA
Val120-Ile201B	(439) AAGTCAACGACAAGAAGTTCAGGGAGGGGCCCTGCA
Val120-Ala204	(433) AAGTCAACGACAAGAAGTTCAGGGAGGGGCCCTGCA
Val120-Ile201	(439) AAGTCAACGACAAGAAGTTCAGGGAGGGGCCCTGCA

Val1120-Thr202	(439)	AAGTGCACAGCAGAAAGTTCACGGCAGCGGCCCTGCA
Lys211-Val120	(445)	AAGTGCACAGCAGAAAGTTCACGGCAGCGGCCCTGCA
Consensus	(481)	AAGTGCACAGCAGAAAGTTCACGGCAGCGGCCCTGCA
	521	560
Leu122-Ser199	(491)	CCACGTCAGCACCGTGCAGTGCACCCACGGCATCCGCC
Val1127-Asn195	(521)	CCACGTCAGCACCGTGCAGTGCACCCACGGCATCCGCC
Val120-Ile201B	(479)	CCACGTCAGCACCGTGCAGTGCACCCACGGCATCCGCC
Val1120-Ala204	(473)	CCACGTCAGCACCGTGCAGTGCACCCACGGCATCCGCC
Val1120-Ile201	(479)	CCACGTCAGCACCGTGCAGTGCACCCACGGCATCCGCC
Val1120-Thr202	(479)	CCACGTCAGCACCGTGCAGTGCACCCACGGCATCCGCC
Lys211-Val120	(485)	CCACGTCAGCACCGTGCAGTGCACCCACGGCATCCGCC
Consensus	(521)	CCACGTCAGCACCGTGCAGTGCACCCACGGCATCCGCC
	561	600
Leu122-Ser199	(531)	CGTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGCC
Val1127-Asn195	(561)	CGTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGCC
Val1120-Ile201B	(519)	CGTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGCC
Val1120-Ala204	(513)	CGTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGCC
Val1120-Ile201	(519)	CGTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGCC
Val1120-Thr202	(519)	CGTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGCC
Lys211-Val120	(525)	CGTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGCC
Consensus	(561)	CGTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGCC
	601	640
Leu122-Ser199	(571)	GAGGAGGGCGTGGTGTATCCGAGGGAGAACTTCACCGACA
Val1127-Asn195	(601)	GAGGAGGGCGTGGTGTATCCGAGGGAGAACTTCACCGACA
Val1120-Ile201B	(559)	GAGGAGGGCGTGGTGTATCCGAGGGAGAACTTCACCGACA
Val1120-Ala204	(553)	GAGGAGGGCGTGGTGTATCCGAGGGAGAACTTCACCGACA
Val1120-Ile201	(559)	GAGGAGGGCGTGGTGTATCCGAGGGAGAACTTCACCGACA
Val1120-Thr202	(559)	GAGGAGGGCGTGGTGTATCCGAGGGAGAACTTCACCGACA
Lys211-Val120	(565)	GAGGAGGGCGTGGTGTATCCGAGGGAGAACTTCACCGACA
Consensus	(601)	GAGGAGGGCGTGGTGTATCCGAGGGAGAACTTCACCGACA
	641	680
Leu122-Ser199	(611)	ACGCCAACGACCATCATCGTCAGCTGAAGGGAGCGTGGA
Val1127-Asn195	(641)	ACGCCAACGACCATCATCGTCAGCTGAAGGGAGCGTGGA
Val1120-Ile201B	(599)	ACGCCAACGACCATCATCGTCAGCTGAAGGGAGCGTGGA
Val1120-Ala204	(593)	ACGCCAACGACCATCATCGTCAGCTGAAGGGAGCGTGGA
Val1120-Ile201	(599)	ACGCCAACGACCATCATCGTCAGCTGAAGGGAGCGTGGA
Val1120-Thr202	(599)	ACGCCAACGACCATCATCGTCAGCTGAAGGGAGCGTGGA
Lys211-Val120	(605)	ACGCCAACGACCATCATCGTCAGCTGAAGGGAGCGTGGA
Consensus	(641)	ACGCCAACGACCATCATCGTCAGCTGAAGGGAGCGTGGA
	681	720
Leu122-Ser199	(651)	GATCAACTGCACCGCCCGAACACAAACACCCGCAAGAGC
Val1127-Asn195	(681)	GATCAACTGCACCGCCCGAACACAAACACCCGCAAGAGC
Val1120-Ile201B	(639)	GATCAACTGCACCGCCCGAACACAAACACCCGCAAGAGC
Val1120-Ala204	(633)	GATCAACTGCACCGCCCGAACACAAACACCCGCAAGAGC
Val1120-Ile201	(639)	GATCAACTGCACCGCCCGAACACAAACACCCGCAAGAGC
Val1120-Thr202	(639)	GATCAACTGCACCGCCCGAACACAAACACCCGCAAGAGC
Lys211-Val120	(645)	GATCAACTGCACCGCCCGAACACAAACACCCGCAAGAGC
Consensus	(681)	GATCAACTGCACCGCCCGAACACAAACACCCGCAAGAGC
	721	760
Leu122-Ser199	(691)	ATCACCATCGGCCCGGGCGGCCCTTCTACGCCAACGGCG
Val1127-Asn195	(721)	ATCACCATCGGCCCGGGCGGCCCTTCTACGCCAACGGCG
Val1120-Ile201B	(679)	ATCACCATCGGCCCGGGCGGCCCTTCTACGCCAACGGCG
Val1120-Ala204	(673)	ATCACCATCGGCCCGGGCGGCCCTTCTACGCCAACGGCG
Val1120-Ile201	(679)	ATCACCATCGGCCCGGGCGGCCCTTCTACGCCAACGGCG
Val1120-Thr202	(679)	ATCACCATCGGCCCGGGCGGCCCTTCTACGCCAACGGCG
Lys211-Val120	(685)	ATCACCATCGGCCCGGGCGGCCCTTCTACGCCAACGGCG
Consensus	(721)	ATCACCATCGGCCCGGGCGGCCCTTCTACGCCAACGGCG

	761	800
Leu122-Ser199	(731) ACATCATCGGCGCATCCGCCAGGGCCACTGCAACATCAG	
Val122-Asn195	(761) ACATCATCGGCGACATCCGCCAGGGCCACTGCAACATCAG	
Val120-Ile201B	(719) ACATCATCGGCGACATCCGCCAGGGCCACTGCAACATCAG	
Val120-Ala204	(713) ACATCATCGGCGACATCCGCCAGGGCCACTGCAACATCAG	
Val120-Ile201	(719) ACATCATCGGCGACATCCGCCAGGGCCACTGCAACATCAG	
Val120-Thr202	(719) ACATCATCGGCGACATCCGCCAGGGCCACTGCAACATCAG	
Lys121-Val200	(725) ACATCATCGGCGACATCCGCCAGGGCCACTGCAACATCAG	
Consensus	(761) ACATCATCGGCGACATCCGCCAGGGCCACTGCAACATCAG	
	801	840
Leu122-Ser199	(771) CGCGGAGAAAGTGGAAACACACCTGAAAGCAGATCGTGACC	
Val122-Asn195	(801) CGCGGAGAAAGTGGAAACACACCTGAAAGCAGATCGTGACC	
Val120-Ile201B	(759) CGCGGAGAAAGTGGAAACACACCTGAAAGCAGATCGTGACC	
Val120-Ala204	(753) CGCGGAGAAAGTGGAAACACACCTGAAAGCAGATCGTGACC	
Val120-Ile201	(759) CGCGGAGAAAGTGGAAACACACCTGAAAGCAGATCGTGACC	
Val120-Thr202	(759) CGCGGAGAAAGTGGAAACACACCTGAAAGCAGATCGTGACC	
Lys121-Val200	(765) CGCGGAGAAAGTGGAAACACACCTGAAAGCAGATCGTGACC	
Consensus	(801) CGCGGAGAAAGTGGAAACACACCTGAAAGCAGATCGTGACC	
	841	880
Leu122-Ser199	(811) AAGCTGCAAGGCCAGTTCGGAAACAGAACATCTGTGTTCA	
Val122-Asn195	(841) AAGCTGCAAGGCCAGTTCGGAAACAGAACATCTGTGTTCA	
Val120-Ile201B	(799) AAGCTGCAAGGCCAGTTCGGAAACAGAACATCTGTGTTCA	
Val120-Ala204	(793) AAGCTGCAAGGCCAGTTCGGAAACAGAACATCTGTGTTCA	
Val120-Ile201	(799) AAGCTGCAAGGCCAGTTCGGAAACAGAACATCTGTGTTCA	
Val120-Thr202	(799) AAGCTGCAAGGCCAGTTCGGAAACAGAACATCTGTGTTCA	
Lys121-Val200	(805) AAGCTGCAAGGCCAGTTCGGAAACAGAACATCTGTGTTCA	
Consensus	(841) AAGCTGCAAGGCCAGTTCGGAAACAGAACATCTGTGTTCA	
	881	920
Leu122-Ser199	(851) AGCAGAGCACGGCGCGCGAGCCCCGAGATCGTATGCACAG	
Val122-Asn195	(881) AGCAGAGCACGGCGCGAGCCCCGAGATCGTATGCACAG	
Val120-Ile201B	(839) AGCAGAGCACGGCGCGAGCCCCGAGATCGTATGCACAG	
Val120-Ala204	(833) AGCAGAGCACGGCGCGAGCCCCGAGATCGTATGCACAG	
Val120-Ile201	(839) AGCAGAGCACGGCGCGAGCCCCGAGATCGTATGCACAG	
Val120-Thr202	(839) AGCAGAGCACGGCGCGAGCCCCGAGATCGTATGCACAG	
Lys121-Val200	(845) AGCAGAGCACGGCGCGAGCCCCGAGATCGTATGCACAG	
Consensus	(881) AGCAGAGCACGGCGCGAGCCCCGAGATCGTATGCACAG	
	921	960
Leu122-Ser199	(891) CTTCAACTGCGGCCGGCGAGTTCTCTACTGCAACAGCACCC	
Val122-Asn195	(921) CTTCAACTGCGGCCGGCGAGTTCTCTACTGCAACAGCACCC	
Val120-Ile201B	(879) CTTCAACTGCGGCCGGCGAGTTCTCTACTGCAACAGCACCC	
Val120-Ala204	(873) CTTCAACTGCGGCCGGCGAGTTCTCTACTGCAACAGCACCC	
Val120-Ile201	(879) CTTCAACTGCGGCCGGCGAGTTCTCTACTGCAACAGCACCC	
Val120-Thr202	(879) CTTCAACTGCGGCCGGCGAGTTCTCTACTGCAACAGCACCC	
Lys121-Val200	(885) CTTCAACTGCGGCCGGCGAGTTCTCTACTGCAACAGCACCC	
Consensus	(921) CTTCAACTGCGGCCGGCGAGTTCTCTACTGCAACAGCACCC	
	961	1000
Leu122-Ser199	(931) CAGCTGTCAACAGCACCTTGGAAACAAACCATCGGCCCCCA	
Val122-Asn195	(961) CAGCTGTCAACAGCACCTTGGAAACAAACCATCGGCCCCCA	
Val120-Ile201B	(919) CAGCTGTCAACAGCACCTTGGAAACAAACCATCGGCCCCCA	
Val120-Ala204	(913) CAGCTGTCAACAGCACCTTGGAAACAAACCATCGGCCCCCA	
Val120-Ile201	(919) CAGCTGTCAACAGCACCTTGGAAACAAACCATCGGCCCCCA	
Val120-Thr202	(919) CAGCTGTCAACAGCACCTTGGAAACAAACCATCGGCCCCCA	
Lys121-Val200	(925) CAGCTGTCAACAGCACCTTGGAAACAAACCATCGGCCCCCA	
Consensus	(961) CAGCTGTCAACAGCACCTTGGAAACAAACCATCGGCCCCCA	
	1001	1040
Leu122-Ser199	(971) ACAACACCAACGGCACCATCACCTGCOCTGGCGCATCAA	
Val122-Asn195	(1001) ACAACACCAACGGCACCATCACCTGCOCTGGCGCATCAA	

FIG. 3D

Val120-Ile201B	(959)	ACAAACACCAACGGCACCATCACCCCTGCCCTGCCGATCAA
Val120-Ala204	(953)	ACAAACACCAACGGCACCATCACCCCTGCCCTGCCGATCAA
Val120-Ile201	(959)	ACAAACACCAACGGCACCATCACCCCTGCCCTGCCGATCAA
Val120-Thr202	(959)	ACAAACACCAACGGCACCATCACCCCTGCCCTGCCGATCAA
Lys121-Val1200	(965)	ACAAACACCAACGGCACCATCACCCCTGCCCTGCCGATCAA
Consensus	(1001)	ACAAACACCAACGGCACCATCACCCCTGCCCTGCCGATCAA 1041 1080
Leu122-Ser199	(1011)	GCAGATCATCACCCGCTGGCAGGAGGTGGGCAAGGCCATG
Val127-Asn195	(1041)	GCAGATCATCACCCGCTGGCAGGAGGTGGGCAAGGCCATG
Val120-Ile201B	(999)	GCAGATCATCACCCGCTGGCAGGAGGTGGGCAAGGCCATG
Val120-Ala204	(993)	GCAGATCATCACCCGCTGGCAGGAGGTGGGCAAGGCCATG
Val120-Ile201	(999)	GCAGATCATCACCCGCTGGCAGGAGGTGGGCAAGGCCATG
Val120-Thr202	(999)	GCAGATCATCACCCGCTGGCAGGAGGTGGGCAAGGCCATG
Lys121-Val1200	(1005)	GCAGATCATCACCCGCTGGCAGGAGGTGGGCAAGGCCATG
Consensus	(1041)	GCAGATCATCACCCGCTGGCAGGAGGTGGGCAAGGCCATG 1081 1120
Leu122-Ser199	(1051)	TACGCCCCCCCCCATCCGGGGCCAGATCCGCTGCAGCAGCA
Val127-Asn195	(1081)	TACGCCCCCCCCCATCCGGGGCCAGATCCGCTGCAGCAGCA
Val120-Ile201B	(1039)	TACGCCCCCCCCCATCCGGGGCCAGATCCGCTGCAGCAGCA
Val120-Ala204	(1033)	TACGCCCCCCCCCATCCGGGGCCAGATCCGCTGCAGCAGCA
Val120-Ile201	(1039)	TACGCCCCCCCCCATCCGGGGCCAGATCCGCTGCAGCAGCA
Val120-Thr202	(1039)	TACGCCCCCCCCCATCCGGGGCCAGATCCGCTGCAGCAGCA
Lys121-Val1200	(1045)	TACGCCCCCCCCCATCCGGGGCCAGATCCGCTGCAGCAGCA
Consensus	(1081)	TACGCCCCCCCCCATCCGGGGCCAGATCCGCTGCAGCAGCA 1121 1160
Leu122-Ser199	(1091)	ACATCACCGGCCCTGCTGCTGACCCCGGACGGCGGCAAGGA
Val127-Asn195	(1121)	ACATCACCGGCCCTGCTGCTGACCCCGGACGGCGGCAAGGA
Val120-Ile201B	(1079)	ACATCACCGGCCCTGCTGCTGACCCCGGACGGCGGCAAGGA
Val120-Ala204	(1073)	ACATCACCGGCCCTGCTGCTGACCCCGGACGGCGGCAAGGA
Val120-Ile201	(1079)	ACATCACCGGCCCTGCTGCTGACCCCGGACGGCGGCAAGGA
Val120-Thr202	(1079)	ACATCACCGGCCCTGCTGCTGACCCCGGACGGCGGCAAGGA
Lys121-Val1200	(1085)	ACATCACCGGCCCTGCTGCTGACCCCGGACGGCGGCAAGGA
Consensus	(1121)	ACATCACCGGCCCTGCTGCTGACCCCGGACGGCGGCAAGGA 1161 1200
Leu122-Ser199	(1131)	GATCAGCAACACCAACCGAGATCTCCGCCCCGGCGCGGC
Val127-Asn195	(1161)	GATCAGCAACACCAACCGAGATCTCCGCCCCGGCGCGGC
Val120-Ile201B	(1119)	GATCAGCAACACCAACCGAGATCTCCGCCCCGGCGCGGC
Val120-Ala204	(1113)	GATCAGCAACACCAACCGAGATCTCCGCCCCGGCGCGGC
Val120-Ile201	(1119)	GATCAGCAACACCAACCGAGATCTCCGCCCCGGCGCGGC
Val120-Thr202	(1119)	GATCAGCAACACCAACCGAGATCTCCGCCCCGGCGCGGC
Lys121-Val1200	(1125)	GATCAGCAACACCAACCGAGATCTCCGCCCCGGCGCGGC
Consensus	(1161)	GATCAGCAACACCAACCGAGATCTCCGCCCCGGCGCGGC 1201 1240
Leu122-Ser199	(1171)	GACATGGCGGACAACTGGCGGAGCTGTAACAGTACA
Val127-Asn195	(1201)	GACATGGCGGACAACTGGCGGAGCTGTAACAGTACA
Val120-Ile201B	(1159)	GACATGGCGGACAACTGGCGGAGCTGTAACAGTACA
Val120-Ala204	(1153)	GACATGGCGGACAACTGGCGGAGCTGTAACAGTACA
Val120-Ile201	(1159)	GACATGGCGGACAACTGGCGGAGCTGTAACAGTACA
Val120-Thr202	(1159)	GACATGGCGGACAACTGGCGGAGCTGTAACAGTACA
Lys121-Val1200	(1165)	GACATGGCGGACAACTGGCGGAGCTGTAACAGTACA
Consensus	(1201)	GACATGGCGGACAACTGGCGGAGCTGTAACAGTACA 1241 1280
Leu122-Ser199	(1211)	AGGTGGTGAAGATCGAGCCCCCTGGCGTGGCCCCCACCAA
Val127-Asn195	(1241)	AGGTGGTGAAGATCGAGCCCCCTGGCGTGGCCCCCACCAA
Val120-Ile201B	(1199)	AGGTGGTGAAGATCGAGCCCCCTGGCGTGGCCCCCACCAA
Val120-Ala204	(1193)	AGGTGGTGAAGATCGAGCCCCCTGGCGTGGCCCCCACCAA
Val120-Ile201	(1199)	AGGTGGTGAAGATCGAGCCCCCTGGCGTGGCCCCCACCAA

Val120-Thr202	(1199)	AGGTGGTGAAGATCGACCCCTGGGGCTGGCCCCACCAA
Lys121-Val1200	(1205)	AGGTGGTGAAGATCGACCCCTGGGGCTGGCCCCACCAA
Consensus	(1241)	AGGTGGTGAAGATCGACCCCTGGGGCTGGCCCCACCAA 1281 1320
Leu122-Ser199	(1251)	GGCCAAGGCCGCGCTGGTGCAGGGCGAGAACCGGCCGTG
Val127-Asn195	(1281)	GGCCAAGGCCGCGCTGGTGCAGGGCGAGAACCGGCCGTG
Val120-Ile201B	(1239)	GGCCAAGGCCGCGCTGGTGCAGGGCGAGAACCGGCCGTG
Val120-Ala204	(1233)	GGCCAAGGCCGCGCTGGTGCAGGGCGAGAACCGGCCGTG
Val120-Ile201	(1239)	GGCCAAGGCCGCGCTGGTGCAGGGCGAGAACCGGCCGTG
Val120-Thr202	(1239)	GGCCAAGGCCGCGCTGGTGCAGGGCGAGAACCGGCCGTG
Lys121-Val1200	(1245)	GGCCAAGGCCGCGCTGGTGCAGGGCGAGAACCGGCCGTG
Consensus	(1281)	GGCCAAGGCCGCGCTGGTGCAGGGCGAGAACCGGCCGTG 1321 1360
Leu122-Ser199	(1291)	ACCTGGGCCGCATGTTCTGGCCTCCGGCGCCCG
Val127-Asn195	(1321)	ACCTGGGCCGCATGTTCTGGCCTCCGGCGCCCG
Val120-Ile201B	(1279)	ACCTGGGCCGCATGTTCTGGCCTCCGGCGCCCG
Val120-Ala204	(1273)	ACCTGGGCCGCATGTTCTGGCCTCCGGCGCCCG
Val120-Ile201	(1279)	ACCTGGGCCGCATGTTCTGGCCTCCGGCGCCCG
Val120-Thr202	(1279)	ACCTGGGCCGCATGTTCTGGCCTCCGGCGCCCG
Lys121-Val1200	(1285)	ACCTGGGCCGCATGTTCTGGCCTCCGGCGCCCG
Consensus	(1321)	ACCTGGGCCGCATGTTCTGGCCTCCGGCGCCCG 1361 1400
Leu122-Ser199	(1331)	GCACCCACCATGGGCCGCCCGACCTGACCCGTGCA
Val127-Asn195	(1361)	GCACCCACCATGGGCCGCCCGACCTGACCCGTGCA
Val120-Ile201B	(1319)	GCACCCACCATGGGCCGCCCGACCTGACCCGTGCA
Val120-Ala204	(1313)	GCACCCACCATGGGCCGCCCGACCTGACCCGTGCA
Val120-Ile201	(1319)	GCACCCACCATGGGCCGCCCGACCTGACCCGTGCA
Val120-Thr202	(1319)	GCACCCACCATGGGCCGCCCGACCTGACCCGTGCA
Lys121-Val1200	(1325)	GCACCCACCATGGGCCGCCCGACCTGACCCGTGCA
Consensus	(1361)	GCACCCACCATGGGCCGCCCGACCTGACCCGTGCA 1401 1440
Leu122-Ser199	(1371)	GCCCCGCCAGCTGCTGGGGCATGTCACAGCAGAAC
Val127-Asn195	(1401)	GCCCCGCCAGCTGCTGGGGCATGTCACAGCAGAAC
Val120-Ile201B	(1359)	GCCCCGCCAGCTGCTGGGGCATGTCACAGCAGAAC
Val120-Ala204	(1353)	GCCCCGCCAGCTGCTGGGGCATGTCACAGCAGAAC
Val120-Ile201	(1359)	GCCCCGCCAGCTGCTGGGGCATGTCACAGCAGAAC
Val120-Thr202	(1359)	GCCCCGCCAGCTGCTGGGGCATGTCACAGCAGAAC
Lys121-Val1200	(1365)	GCCCCGCCAGCTGCTGGGGCATGTCACAGCAGAAC
Consensus	(1401)	GCCCCGCCAGCTGCTGGGGCATGTCACAGCAGAAC 1441 1480
Leu122-Ser199	(1411)	ACCTGGTGGGCCGCATCGAGGCCACAGCACCTGCTG
Val127-Asn195	(1441)	ACCTGGTGGGCCGCATCGAGGCCACAGCACCTGCTG
Val120-Ile201B	(1399)	ACCTGGTGGGCCGCATCGAGGCCACAGCACCTGCTG
Val120-Ala204	(1393)	ACCTGGTGGGCCGCATCGAGGCCACAGCACCTGCTG
Val120-Ile201	(1399)	ACCTGGTGGGCCGCATCGAGGCCACAGCACCTGCTG
Val120-Thr202	(1399)	ACCTGGTGGGCCGCATCGAGGCCACAGCACCTGCTG
Lys121-Val1200	(1405)	ACCTGGTGGGCCGCATCGAGGCCACAGCACCTGCTG
Consensus	(1441)	ACCTGGTGGGCCGCATCGAGGCCACAGCACCTGCTG 1481 1520
Leu122-Ser199	(1451)	AGCTGACCGFTGGGGCATCGAGGCCACAGCACCTGCTG
Val127-Asn195	(1481)	AGCTGACCGFTGGGGCATCGAGGCCACAGCACCTGCTG
Val120-Ile201B	(1439)	AGCTGACCGFTGGGGCATCGAGGCCACAGCACCTGCTG
Val120-Ala204	(1433)	AGCTGACCGFTGGGGCATCGAGGCCACAGCACCTGCTG
Val120-Ile201	(1439)	AGCTGACCGFTGGGGCATCGAGGCCACAGCACCTGCTG
Val120-Thr202	(1439)	AGCTGACCGFTGGGGCATCGAGGCCACAGCACCTGCTG
Lys121-Val1200	(1445)	AGCTGACCGFTGGGGCATCGAGGCCACAGCACCTGCTG
Consensus	(1481)	AGCTGACCGFTGGGGCATCGAGGCCACAGCACCTGCTG

		1521		1560
Leu122-Ser199	(1491)	GCTGGCCGTGGAGCGCTACCTGAAAGGACCAAGCAGCTGCTG		
Val127-Asn195	(1521)	GCTGGCCGTGGAGCGCTACCTGAAAGGACCAAGCAGCTGCTG		
Val120-1le201B	(1479)	GCTGGCCGTGGAGCGCTACCTGAAAGGACCAAGCAGCTGCTG		
Val120-Ala204	(1473)	GCTGGCCGTGGAGCGCTACCTGAAAGGACCAAGCAGCTGCTG		
Val120-1le201	(1479)	GCTGGCCGTGGAGCGCTACCTGAAAGGACCAAGCAGCTGCTG		
Val120-Thr202	(1479)	GCTGGCCGTGGAGCGCTACCTGAAAGGACCAAGCAGCTGCTG		
Lys121-Val200	(1485)	GCTGGCCGTGGAGCGCTACCTGAAAGGACCAAGCAGCTGCTG		
Consensus	(1521)	GCTGGCCGTGGAGCGCTACCTGAAAGGACCAAGCAGCTGCTG		
		1561		1600
Leu122-Ser199	(1531)	GGCATCTGGGCTGCAGGCGAAAGCTGATCTGCACCAACCG		
Val127-Asn195	(1561)	GGCATCTGGGCTGCAGGCGAAAGCTGATCTGCACCAACCG		
Val120-1le201B	(1519)	GGCATCTGGGCTGCAGGCGAAAGCTGATCTGCACCAACCG		
Val120-Ala204	(1513)	GGCATCTGGGCTGCAGGCGAAAGCTGATCTGCACCAACCG		
Val120-1le201	(1519)	GGCATCTGGGCTGCAGGCGAAAGCTGATCTGCACCAACCG		
Val120-Thr202	(1519)	GGCATCTGGGCTGCAGGCGAAAGCTGATCTGCACCAACCG		
Lys121-Val200	(1525)	GGCATCTGGGCTGCAGGCGAAAGCTGATCTGCACCAACCG		
Consensus	(1561)	GGCATCTGGGCTGCAGGCGAAAGCTGATCTGCACCAACCG		
		1601		1640
Leu122-Ser199	(1571)	CCGTGGCTCTGGAAACCCAGCTGGAGCAACAAAGGCTTGA		
Val127-Asn195	(1601)	CCGTGGCTCTGGAAACCCAGCTGGAGCAACAAAGGCTTGA		
Val120-1le201B	(1559)	CCGTGGCTCTGGAAACCCAGCTGGAGCAACAAAGGCTTGA		
Val120-Ala204	(1553)	CCGTGGCTCTGGAAACCCAGCTGGAGCAACAAAGGCTTGA		
Val120-1le201	(1559)	CCGTGGCTCTGGAAACCCAGCTGGAGCAACAAAGGCTTGA		
Val120-Thr202	(1559)	CCGTGGCTCTGGAAACCCAGCTGGAGCAACAAAGGCTTGA		
Lys121-Val200	(1565)	CCGTGGCTCTGGAAACCCAGCTGGAGCAACAAAGGCTTGA		
Consensus	(1601)	CCGTGGCTCTGGAAACCCAGCTGGAGCAACAAAGGCTTGA		
		1641		1680
Leu122-Ser199	(1611)	CCAGATCTGGAAACATGACCTGGATGGAGTGGAGCTGGC		
Val127-Asn195	(1641)	CCAGATCTGGAAACATGACCTGGATGGAGTGGAGCTGGC		
Val120-1le201B	(1599)	CCAGATCTGGAAACATGACCTGGATGGAGTGGAGCTGGC		
Val120-Ala204	(1593)	CCAGATCTGGAAACATGACCTGGATGGAGTGGAGCTGGC		
Val120-1le201	(1599)	CCAGATCTGGAAACATGACCTGGATGGAGTGGAGCTGGC		
Val120-Thr202	(1599)	CCAGATCTGGAAACATGACCTGGATGGAGTGGAGCTGGC		
Lys121-Val200	(1605)	CCAGATCTGGAAACATGACCTGGATGGAGTGGAGCTGGC		
Consensus	(1641)	CCAGATCTGGAAACATGACCTGGATGGAGTGGAGCTGGC		
		1681		1720
Leu122-Ser199	(1651)	GAGATCGAAACTACACCAACCTGATCTACACCCCTGATCG		
Val127-Asn195	(1681)	GAGATCGAAACTACACCAACCTGATCTACACCCCTGATCG		
Val120-1le201B	(1639)	GAGATCGAAACTACACCAACCTGATCTACACCCCTGATCG		
Val120-Ala204	(1633)	GAGATCGAAACTACACCAACCTGATCTACACCCCTGATCG		
Val120-1le201	(1639)	GAGATCGAAACTACACCAACCTGATCTACACCCCTGATCG		
Val120-Thr202	(1639)	GAGATCGAAACTACACCAACCTGATCTACACCCCTGATCG		
Lys121-Val200	(1645)	GAGATCGAAACTACACCAACCTGATCTACACCCCTGATCG		
Consensus	(1681)	GAGATCGAAACTACACCAACCTGATCTACACCCCTGATCG		
		1721		1760
Leu122-Ser199	(1691)	AGGAGAGCCAGAACCCAGCAGGAGAACGAGCAGCAGGAGCT		
Val127-Asn195	(1721)	AGGAGAGCCAGAACCCAGCAGGAGAACGAGCAGCAGGAGCT		
Val120-1le201B	(1679)	AGGAGAGCCAGAACCCAGCAGGAGAACGAGCAGCAGGAGCT		
Val120-Ala204	(1673)	AGGAGAGCCAGAACCCAGCAGGAGAACGAGCAGCAGGAGCT		
Val120-1le201	(1679)	AGGAGAGCCAGAACCCAGCAGGAGAACGAGCAGCAGGAGCT		
Val120-Thr202	(1679)	AGGAGAGCCAGAACCCAGCAGGAGAACGAGCAGCAGGAGCT		
Lys121-Val200	(1685)	AGGAGAGCCAGAACCCAGCAGGAGAACGAGCAGCAGGAGCT		
Consensus	(1721)	AGGAGAGCCAGAACCCAGCAGGAGAACGAGCAGCAGGAGCT		
		1761		1800
Leu122-Ser199	(1731)	GCTGGAGCTGGACAAAGTGGCCAGCCTGTGGAAGTGGTTC		
Val127-Asn195	(1761)	GCTGGAGCTGGACAAAGTGGCCAGCCTGTGGAAGTGGTTC		

FIG. 3G

Val1120-Ile2018	(1719)	GCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTC
Val1120-Ala204	(1713)	GCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTC
Val1120-Ile201	(1719)	GCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTC
Val1120-Thr202	(1719)	GCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTC
Lys121-Val1200	(1725)	GCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTC
Consensus	(1761)	GCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTC 1801 1840
Leu1122-Ser199	(1771)	GACATCAGCAAGTGGCTGGTACATCAAGATCTTCATCA
Val1127-Asn195	(1801)	GACATCAGCAAGTGGCTGGTACATCAAGATCTTCATCA
Val1120-Ile2018	(1759)	GACATCAGCAAGTGGCTGGTACATCAAGATCTTCATCA
Val1120-Ala204	(1753)	GACATCAGCAAGTGGCTGGTACATCAAGATCTTCATCA
Val1120-Ile201	(1759)	GACATCAGCAAGTGGCTGGTACATCAAGATCTTCATCA
Val1120-Thr202	(1759)	GACATCAGCAAGTGGCTGGTACATCAAGATCTTCATCA
Lys121-Val1200	(1765)	GACATCAGCAAGTGGCTGGTACATCAAGATCTTCATCA
Consensus	(1801)	GACATCAGCAAGTGGCTGGTACATCAAGATCTTCATCA 1841 1880
Leu1122-Ser199	(1811)	TGATCGTGGCGGCCCTGGTGGGCTGCGCATCGTGTTCAC
Val1127-Asn195	(1841)	TGATCGTGGCGGCCCTGGTGGGCTGCGCATCGTGTTCAC
Val1120-Ile2018	(1799)	TGATCGTGGCGGCCCTGGTGGGCTGCGCATCGTGTTCAC
Val1120-Ala204	(1793)	TGATCGTGGCGGCCCTGGTGGGCTGCGCATCGTGTTCAC
Val1120-Ile201	(1793)	TGATCGTGGCGGCCCTGGTGGGCTGCGCATCGTGTTCAC
Val1120-Thr202	(1799)	TGATCGTGGCGGCCCTGGTGGGCTGCGCATCGTGTTCAC
Lys121-Val1200	(1805)	TGATCGTGGCGGCCCTGGTGGGCTGCGCATCGTGTTCAC
Consensus	(1841)	TGATCGTGGCGGCCCTGGTGGGCTGCGCATCGTGTTCAC 1881 1920
Leu1122-Ser199	(1851)	CGCTCTGAGGATCGTGACCCCTGGCCAGGGCTACAGC
Val1127-Asn195	(1881)	CGCTCTGAGGATCGTGACCCCTGGCCAGGGCTACAGC
Val1120-Ile2018	(1839)	CGCTCTGAGGATCGTGACCCCTGGCCAGGGCTACAGC
Val1120-Ala204	(1833)	CGCTCTGAGGATCGTGACCCCTGGCCAGGGCTACAGC
Val1120-Ile201	(1839)	CGCTCTGAGGATCGTGACCCCTGGCCAGGGCTACAGC
Val1120-Thr202	(1839)	CGCTCTGAGGATCGTGACCCCTGGCCAGGGCTACAGC
Lys121-Val1200	(1845)	CGCTCTGAGGATCGTGACCCCTGGCCAGGGCTACAGC
Consensus	(1881)	CGCTCTGAGGATCGTGACCCCTGGCCAGGGCTACAGC 1921 1960
Leu1122-Ser199	(1891)	CCCCCTGAGCTTCAGACCCCTTCCCCGCCCGGCC
Val1127-Asn195	(1921)	CCCCCTGAGCTTCAGACCCCTTCCCCGCCCGGCC
Val1120-Ile2018	(1879)	CCCCCTGAGCTTCAGACCCCTTCCCCGCCCGGCC
Val1120-Ala204	(1873)	CCCCCTGAGCTTCAGACCCCTTCCCCGCCCGGCC
Val1120-Ile201	(1879)	CCCCCTGAGCTTCAGACCCCTTCCCCGCCCGGCC
Val1120-Thr202	(1879)	CCCCCTGAGCTTCAGACCCCTTCCCCGCCCGGCC
Lys121-Val1200	(1885)	CCCCCTGAGCTTCAGACCCCTTCCCCGCCCGGCC
Consensus	(1921)	CCCCCTGAGCTTCAGACCCCTTCCCCGCCCGGCC 1961 2000
Leu1122-Ser199	(1931)	CGACACGGCGCCCGAGGGCATCGAGGAGGAGGCCGGCG
Val1127-Asn195	(1961)	CGACACGGCGCCCGAGGGCATCGAGGAGGAGGCCGG
Val1120-Ile2018	(1919)	CGACACGGCGCCCGAGGGCATCGAGGAGGAGGCCGG
Val1120-Ala204	(1913)	CGACACGGCGCCCGAGGGCATCGAGGAGGAGGCCGG
Val1120-Ile201	(1919)	CGACACGGCGCCCGAGGGCATCGAGGAGGAGGCCGG
Val1120-Thr202	(1919)	CGACACGGCGCCCGAGGGCATCGAGGAGGAGGCCGG
Lys121-Val1200	(1925)	CGACACGGCGCCCGAGGGCATCGAGGAGGAGGCCGG
Consensus	(1961)	CGACACGGCGCCCGAGGGCATCGAGGAGGAGGCCGG 2001 2040
Leu1122-Ser199	(1971)	CGACACGGCGACCGCGAGCAGCCCCCTGGTGCACGGCTG
Val1127-Asn195	(2001)	CGACACGGCGACCGCGAGCAGCCCCCTGGTGCACGGCTG
Val1120-Ile2018	(1959)	CGACACGGCGACCGCGAGCAGCCCCCTGGTGCACGGCTG
Val1120-Ala204	(1953)	CGACACGGCGACCGCGAGCAGCCCCCTGGTGCACGGCTG
Val1120-Ile201	(1959)	CGACACGGCGACCGCGAGCAGCCCCCTGGTGCACGGCTG

Val120-Thr202	(1959)	CGACCCGCGACCGCAGCAGCCCCCTGGTGACGGCCTGCTG
Lys121-Val200	(1965)	CGACCCGCGACCGCAGCAGCCCCCTGGTGACGGCCTGCTG
Consensus	(2001)	CGACCCGCGACCGCAGCAGCCCCCTGGTGACGGCCTGCTG
	2041	2080
Leu122-Ser199	(2011)	GCCCTGACTGGGACGACCTCGCGAGCCTTGCCCTGTTCA
Val122-Asn195	(2041)	GCCCTGACTGGGACGACCTCGCGAGCCTTGCCCTGTTCA
Val120-Ile201B	(1999)	GCCCTGAATCTGGGACGACCTCGCGAGCCTTGCCCTGTTCA
Val120-Ala204	(1993)	GCCCTGAATCTGGGACGACCTCGCGAGCCTTGCCCTGTTCA
Val120-Ile201	(1999)	GCCCTGAATCTGGGACGACCTCGCGAGCCTTGCCCTGTTCA
Val120-Thr202	(1999)	GCCCTGAATCTGGGACGACCTCGCGAGCCTTGCCCTGTTCA
Lys121-Val200	(2005)	GCCCTGAATCTGGGACGACCTCGCGAGCCTTGCCCTGTTCA
Consensus	(2041)	GCCCTGAATCTGGGACGACCTCGCGAGCCTTGCCCTGTTCA
	2081	2120
Leu122-Ser199	(2051)	GCTACACCCGCTCGCGRCCTGATCTGTATCCGGCCCG
Val122-Asn195	(2081)	GCTACACCCGCTCGCGRCCTGATCTGTATCCGGCCCG
Val120-Ile201B	(2039)	GCTACACCCGCTCGCGRCCTGATCTGTATCCGGCCCG
Val120-Ala204	(2033)	GCTACACCCGCTCGCGRCCTGATCTGTATCCGGCCCG
Val120-Ile201	(2039)	GCTACACCCGCTCGCGRCCTGATCTGTATCCGGCCCG
Val120-Thr202	(2039)	GCTACACCCGCTCGCGRCCTGATCTGTATCCGGCCCG
Lys121-Val200	(2045)	GCTACACCCGCTCGCGRCCTGATCTGTATCCGGCCCG
Consensus	(2081)	GCTACACCCGCTCGCGRCCTGATCTGTATCCGGCCCG
	2121	2160
Leu122-Ser199	(2091)	CATCTGGAGCTGCTGGGCGCCGGCTGGAGGGCCCTG
Val122-Asn195	(2121)	CATCTGGAGCTGCTGGGCGCCGGCTGGAGGGCCCTG
Val120-Ile201B	(2079)	CATCTGGAGCTGCTGGGCGCCGGCTGGAGGGCCCTG
Val120-Ala204	(2073)	CATCTGGAGCTGCTGGGCGCCGGCTGGAGGGCCCTG
Val120-Ile201	(2079)	CATCTGGAGCTGCTGGGCGCCGGCTGGAGGGCCCTG
Val120-Thr202	(2079)	CATCTGGAGCTGCTGGGCGCCGGCTGGAGGGCCCTG
Lys121-Val200	(2085)	CATCTGGAGCTGCTGGGCGCCGGCTGGAGGGCCCTG
Consensus	(2121)	CATCTGGAGCTGCTGGGCGCCGGCTGGAGGGCCCTG
	2161	2200
Leu122-Ser199	(2131)	AACTACTGGSCAACCTGCTGCCAGTCTGGATCAGGAGC
Val122-Asn195	(2161)	AACTACTGGSCAACCTGCTGCCAGTCTGGATCAGGAGC
Val120-Ile201B	(2119)	AACTACTGGSCAACCTGCTGCCAGTCTGGATCAGGAGC
Val120-Ala204	(2113)	AACTACTGGSCAACCTGCTGCCAGTCTGGATCAGGAGC
Val120-Ile201	(2119)	AACTACTGGSCAACCTGCTGCCAGTCTGGATCAGGAGC
Val120-Thr202	(2119)	AACTACTGGSCAACCTGCTGCCAGTCTGGATCAGGAGC
Lys121-Val200	(2125)	AACTACTGGSCAACCTGCTGCCAGTCTGGATCAGGAGC
Consensus	(2161)	AACTACTGGSCAACCTGCTGCCAGTCTGGATCAGGAGC
	2201	2240
Leu122-Ser199	(2171)	TGAAAGACAGCGCTGAGCTTGTCGCCCATGCCAT
Val122-Asn195	(2201)	TGAAAGACAGCGCTGAGCTTGTCGCCCATGCCAT
Val120-Ile201B	(2159)	TGAAAGACAGCGCTGAGCTTGTCGCCCATGCCAT
Val120-Ala204	(2153)	TGAAAGACAGCGCTGAGCTTGTCGCCCATGCCAT
Val120-Ile201	(2159)	TGAAAGACAGCGCTGAGCTTGTCGCCCATGCCAT
Val120-Thr202	(2159)	TGAAAGACAGCGCTGAGCTTGTCGCCCATGCCAT
Lys121-Val200	(2165)	TGAAAGACAGCGCTGAGCTTGTCGCCCATGCCAT
Consensus	(2201)	TGAAAGACAGCGCTGAGCTTGTCGCCCATGCCAT
	2241	2280
Leu122-Ser199	(2211)	CGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCC
Val122-Asn195	(2241)	CGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCC
Val120-Ile201B	(2199)	CGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCC
Val120-Ala204	(2193)	CGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCC
Val120-Ile201	(2199)	CGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCC
Val120-Thr202	(2199)	CGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCC
Lys121-Val200	(2205)	CGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCC
Consensus	(2241)	CGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCC

FIG. 3I

	2281	2320
Leu122-Ser199	(2251) CAGCGCATCGGCCGCGCCTTCTGACATCCCCCGCCGA	
Val127-Asn195	(2281) CAGCGCATCGGCCGCGCCTTCTGACATCCCCCGCCGA	
Val120-Ile201B	(2239) CAGCGCATCGGCCGCGCCTTCTGACATCCCCCGCCGA	
Val120-Ala204	(2233) CAGCGCATCGGCCGCGCCTTCTGACATCCCCCGCCGA	
Val120-Ile201	(2239) CAGCGCATCGGCCGCGCCTTCTGACATCCCCCGCCGA	
Val120-Thr202	(2239) CAGCGCATCGGCCGCGCCTTCTGACATCCCCCGCCGA	
Lys121-Val200	(2245) CAGCGCATCGGCCGCGCCTTCTGACATCCCCCGCCGA	
Consensus	(2281) CAGCGCATCGGCCGCGCCTTCTGACATCCCCCGCCGA	
	2321	2360
Leu122-Ser199	(2291) TCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAACTCGAGCG	
Val127-Asn195	(2321) TCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAACTCGAG--	
Val120-Ile201B	(2279) TCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAACTCGAGCG	
Val120-Ala204	(2273) TCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAACTCGAG--	
Val120-Ile201	(2279) TCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAACTCGAG--	
Val120-Thr202	(2279) TCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAACTCGAG--	
Lys121-Val200	(2285) TCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAACTCGAGCG	
Consensus	(2321) TCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAACTCGAG	
	2361	
Leu122-Ser199	(2331) TGCT	
Val127-Asn195	(2359) ----	
Val120-Ile201B	(2319) TGCT	
Val120-Ala204	(2311) ----	
Val120-Ile201	(2317) ----	
Val120-Thr202	(2317) ----	
Lys121-Val200	(2325) TGCT	
Consensus	(2361)	

FIG. 3J

	1	40
Ile424-Ala433	(1)	
Trp427-Gly431	(1)	
Gln422-Tyr435B	(1)	
Arg426-Gly431	(1)	
Ile423-Met434	(1)	
Gln422-Tyr435	(1)	
Arg426-Lys432	(1)	
Arg426-Gly431B	(1)	
Asn425-Lys432	(1)	
Consensus	(1) GAATTCGCCACCATGGATGCAATGAAGAGGGGTCTGCT	
	41	80
Ile424-Ala433	(41)	
Trp427-Gly431	(41)	
Gln422-Tyr435B	(41)	
Arg426-Gly431	(41)	
Ile423-Met434	(41)	
Gln422-Tyr435	(41)	
Arg426-Lys432	(41)	
Arg426-Gly431B	(41)	
Asn425-Lys432	(41)	
Consensus	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTGCCAG	
	81	120
Ile424-Ala433	(81)	
Trp427-Gly431	(81)	
Gln422-Tyr435B	(81)	
Arg426-Gly431	(81)	
Ile423-Met434	(81)	
Gln422-Tyr435	(81)	
Arg426-Lys432	(81)	
Arg426-Gly431B	(81)	
Asn425-Lys432	(81)	
Consensus	(81) CGCGGTGGAGAAGCTGTGGTGCCGTGTACTACGGCGTG	
	121	160
Ile424-Ala433	(121)	
Trp427-Gly431	(121)	
Gln422-Tyr435B	(121)	
Arg426-Gly431	(121)	
Ile423-Met434	(121)	
Gln422-Tyr435	(121)	
Arg426-Lys432	(121)	
Arg426-Gly431B	(121)	
Asn425-Lys432	(121)	
Consensus	(121) CCCGTGTTGGAGGAGGCCACCAACCTGTCTGGCCA	
	161	200
Ile424-Ala433	(161)	
Trp427-Gly431	(161)	
Gln422-Tyr435B	(161)	
Arg426-Gly431	(161)	
Ile423-Met434	(161)	
Gln422-Tyr435	(161)	
Arg426-Lys432	(161)	
Arg426-Gly431B	(161)	
Asn425-Lys432	(161)	
Consensus	(161) GCGACCCCAAGGCCCTACGACACCGAGGTGCACACGTTGT	
	201	240
Ile424-Ala433	(201)	

FIG. 4A

Trp427-Gly431	(201)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Gln422-Tyr435B	(201)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Arg426-Gly431	(201)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Ile423-Met434	(201)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Gln422-Tyr435	(201)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Arg426-Lys432	(201)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Arg426-Gly431B	(201)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Asn425-Lys432	(201)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Consensus	(201)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
	241	280
Ile424-Ala433	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Trp427-Gly431	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Gln422-Tyr435B	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Arg426-Gly431	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Ile423-Met434	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Gln422-Tyr435	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Arg426-Lys432	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Arg426-Gly431B	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Asn425-Lys432	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Consensus	(241)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
	281	320
Ile424-Ala433	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
Trp427-Gly431	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
Gln422-Tyr435B	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
Arg426-Gly431	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
Ile423-Met434	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
Gln422-Tyr435	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
Arg426-Lys432	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
Arg426-Gly431B	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
Asn425-Lys432	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
Consensus	(281)	GGAGAACAAACATGGTGGAGCAGATGCACGGGACATCAT
	321	360
Ile424-Ala433	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Trp427-Gly431	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Gln422-Tyr435B	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Arg426-Gly431	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Ile423-Met434	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Gln422-Tyr435	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Arg426-Lys432	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Arg426-Gly431B	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Asn425-Lys432	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
Consensus	(321)	GGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGT
	361	400
Ile424-Ala433	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Trp427-Gly431	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Gln422-Tyr435B	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Arg426-Gly431	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Ile423-Met434	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Gln422-Tyr435	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Arg426-Lys432	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Arg426-Gly431B	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Asn425-Lys432	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Consensus	(361)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
	361	400
Ile424-Ala433	(401)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Trp427-Gly431	(401)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG
Gln422-Tyr435B	(401)	GGCCACCCACGCCCTGCGTGCCCCACCGACCCCCACCCCCAG

FIG. 4B

Arg426-Gly431	(401)	deipeptidyl peptidase 1 substrate	401
Ile423-Met434	(401)	deipeptidyl peptidase 1 substrate	401
Gln422-Tyr435B	(401)	deipeptidyl peptidase 1 substrate	401
Arg426-Lys432	(401)	deipeptidyl peptidase 1 substrate	401
Arg426-Gly431B	(401)	deipeptidyl peptidase 1 substrate	401
Asn425-Lys432	(401)	deipeptidyl peptidase 1 substrate	401
Consensus	(401)	ACGCCACCAACACCAAGAGCAGCAACTGCGAAGGAGATGGAA	480
	441		
Ile424-Ala433	(441)	deipeptidyl peptidase 1 substrate	441
Trp427-Gly431	(441)	deipeptidyl peptidase 1 substrate	441
Gln422-Tyr435B	(441)	deipeptidyl peptidase 1 substrate	441
Arg426-Gly431	(441)	deipeptidyl peptidase 1 substrate	441
Ile423-Met434	(441)	deipeptidyl peptidase 1 substrate	441
Gln422-Tyr435	(441)	deipeptidyl peptidase 1 substrate	441
Arg426-Lys432	(441)	deipeptidyl peptidase 1 substrate	441
Arg426-Gly431B	(441)	deipeptidyl peptidase 1 substrate	441
Asn425-Lys432	(441)	deipeptidyl peptidase 1 substrate	441
Consensus	(441)	CCGGCGGCGAGATCAAGAACTGCAGCTTCAGGTGACCACCA	520
	481		
Ile424-Ala433	(481)	deipeptidyl peptidase 1 substrate	481
Trp427-Gly431	(481)	deipeptidyl peptidase 1 substrate	481
Gln422-Tyr435B	(481)	deipeptidyl peptidase 1 substrate	481
Arg426-Gly431	(481)	deipeptidyl peptidase 1 substrate	481
Ile423-Met434	(481)	deipeptidyl peptidase 1 substrate	481
Gln422-Tyr435	(481)	deipeptidyl peptidase 1 substrate	481
Arg426-Lys432	(481)	deipeptidyl peptidase 1 substrate	481
Arg426-Gly431B	(481)	deipeptidyl peptidase 1 substrate	481
Asn425-Lys432	(481)	deipeptidyl peptidase 1 substrate	481
Consensus	(481)	AGCATCGCAACAAAGATGCAGAGGAGTAGCCCCCTGTTCT	560
	521		
Ile424-Ala433	(521)	deipeptidyl peptidase 1 substrate	521
Trp427-Gly431	(521)	deipeptidyl peptidase 1 substrate	521
Gln422-Tyr435B	(521)	deipeptidyl peptidase 1 substrate	521
Arg426-Gly431	(521)	deipeptidyl peptidase 1 substrate	521
Ile423-Met434	(521)	deipeptidyl peptidase 1 substrate	521
Gln422-Tyr435	(521)	deipeptidyl peptidase 1 substrate	521
Arg426-Lys432	(521)	deipeptidyl peptidase 1 substrate	521
Arg426-Gly431B	(521)	deipeptidyl peptidase 1 substrate	521
Asn425-Lys432	(521)	deipeptidyl peptidase 1 substrate	521
Consensus	(521)	ACAAAGCTGGACGTGGTGCCTCATGCACRGCACACRCCAG	600
	561		
Ile424-Ala433	(561)	deipeptidyl peptidase 1 substrate	561
Trp427-Gly431	(561)	deipeptidyl peptidase 1 substrate	561
Gln422-Tyr435B	(561)	deipeptidyl peptidase 1 substrate	561
Arg426-Gly431	(561)	deipeptidyl peptidase 1 substrate	561
Ile423-Met434	(561)	deipeptidyl peptidase 1 substrate	561
Gln422-Tyr435	(561)	deipeptidyl peptidase 1 substrate	561
Arg426-Lys432	(561)	deipeptidyl peptidase 1 substrate	561
Arg426-Gly431B	(561)	deipeptidyl peptidase 1 substrate	561
Asn425-Lys432	(561)	deipeptidyl peptidase 1 substrate	561
Consensus	(561)	CTACAAGCTGATCAACTGCAACACCCAGCGTGTACCCAGG	640
	601		
Ile424-Ala433	(601)	deipeptidyl peptidase 1 substrate	601
Trp427-Gly431	(601)	deipeptidyl peptidase 1 substrate	601
Gln422-Tyr435B	(601)	deipeptidyl peptidase 1 substrate	601
Arg426-Gly431	(601)	deipeptidyl peptidase 1 substrate	601
Ile423-Met434	(601)	deipeptidyl peptidase 1 substrate	601

FIG. 4G

Gln422-Tyr435	(601)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Lys432	(601)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431B	(601)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Asn425-Lys432	(601)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Consensus	(601)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
	641	680
Ile424-Ala433	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Trp427-Gly431	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435B	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Ile423-Met434	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Lys432	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431B	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Asn425-Lys432	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Consensus	(641)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
	681	720
Ile424-Ala433	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Trp427-Gly431	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435B	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Ile423-Met434	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Lys432	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431B	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Asn425-Lys432	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Consensus	(681)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
	721	760
Ile424-Ala433	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Trp427-Gly431	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435B	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Ile423-Met434	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Lys432	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431B	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Asn425-Lys432	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Consensus	(721)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
	761	800
Ile424-Ala433	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Trp427-Gly431	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435B	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Ile423-Met434	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Lys432	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431B	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Asn425-Lys432	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Consensus	(761)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
	801	840
Ile424-Ala433	(801)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Trp427-Gly431	(801)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435B	(801)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Gly431	(801)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Ile423-Met434	(801)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Gln422-Tyr435	(801)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT
Arg426-Lys432	(801)	ATGCTGCGCCCAAGGTGAGCTTCGAGCCATCCCCATCCACT

FIG. 4D

Arg426-Gly431B	(801)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Asn425-Lys432	(801)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Consensus	(801)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
	841	880
Ile424-Ala433	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
Trp427-Gly431	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
Gln422-Tyr435B	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
Arg426-Gly431	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
Ile423-Met434	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
Gln422-Tyr435	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
Arg426-Lys432	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
Arg426-Gly431B	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
Asn425-Lys432	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
Consensus	(841)	ATCATCGTCAGCTGAGGGAGAGCTGGAGATCAACTGCA
	881	920
Ile424-Ala433	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Trp427-Gly431	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Gln422-Tyr435B	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Arg426-Gly431	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Ile423-Met434	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Gln422-Tyr435	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Arg426-Lys432	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Arg426-Gly431B	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Asn425-Lys432	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Consensus	(881)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
	921	960
Ile424-Ala433	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Trp427-Gly431	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Gln422-Tyr435B	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Arg426-Gly431	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Ile423-Met434	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Gln422-Tyr435	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Arg426-Lys432	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Arg426-Gly431B	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Asn425-Lys432	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
Consensus	(921)	CCCCGGCCCAACCAACACCCCCCAAGAGCATCACATCGG
	961	1000
Ile424-Ala433	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
Trp427-Gly431	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
Gln422-Tyr435B	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
Arg426-Gly431	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
Ile423-Met434	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
Gln422-Tyr435	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
Arg426-Lys432	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
Arg426-Gly431B	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
Asn425-Lys432	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
Consensus	(961)	GACATCGCCAGGCCACTGCAACATCAGCGCGAGAACT
	1001	1040
Ile424-Ala433	(1001)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Trp427-Gly431	(1001)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Gln422-Tyr435B	(1001)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Arg426-Gly431	(1001)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Ile423-Met434	(1001)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Gln422-Tyr435	(1001)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Arg426-Lys432	(1001)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Arg426-Gly431B	(1001)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC
Asn425-Lys432	(1001)	GGTGGATCCGGCAGCGAGAACTTCACCGACAACGCCAACCC

FIG. 4E

Consensus	(1001)	GGAACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
	1041	1080
Ile424-Ala433	(1041)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Trp427-Gly431	(1041)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435B	(1041)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431	(1041)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Ile423-Met434	(1041)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435	(1041)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Lys432	(1041)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431B	(1041)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Asn425-Lys432	(1041)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Consensus	(1041)	CCGGTCGCCAACAAAGACCCTCGTGTCAAGCAGAGCAGC
	1081	1120
Ile424-Ala433	(1081)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Trp427-Gly431	(1081)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435B	(1081)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431	(1081)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Ile423-Met434	(1081)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435	(1081)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Lys432	(1081)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431B	(1081)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Asn425-Lys432	(1081)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Consensus	(1081)	GGCGGGCAGCCCCGAGATCGTGATGCACAGCTTAACTCGGC
	1121	1160
Ile424-Ala433	(1121)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Trp427-Gly431	(1121)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435B	(1121)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431	(1121)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Ile423-Met434	(1121)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435	(1121)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Lys432	(1121)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431B	(1121)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Asn425-Lys432	(1121)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Consensus	(1121)	GGCGGGCAGTTCTCTACTGCAACAGCACCCAGCTGTCAA
	1161	1200
Ile424-Ala433	(1161)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Trp427-Gly431	(1161)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435B	(1161)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431	(1161)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Ile423-Met434	(1161)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435	(1161)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Lys432	(1161)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431B	(1161)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Asn425-Lys432	(1161)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Consensus	(1161)	CAGCACCTGGAACACCCATCGGCCCCAACACCAAC
	1201	1240
Ile424-Ala433	(1201)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Trp427-Gly431	(1201)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435B	(1201)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431	(1201)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Ile423-Met434	(1201)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Gln422-Tyr435	(1201)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Lys432	(1201)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Arg426-Gly431B	(1201)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Asn425-Lys432	(1201)	GGGACACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGC
Consensus	(1201)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCA
	1241	1280

FIG. 4F

21 / 65

FIG. 4G

FIG. 4H

Ile423-Met434	(1623)	CGAGGCCCCAGCAGCACCTGCTGCAGCTGACCGTGTTGGGGC
Gln422-Tyr435	(1617)	
Arg426-Lys432	(1641)	
Arg426-Gly431B	(1641)	
Asn425-Lys432	(1635)	
Consensus	(1641)	
	1681	1720
Ile424-Ala433	(1669)	
Trp427-Gly431	(1681)	
Gln422-Tyr435B	(1657)	
Arg426-Gly431	(1681)	
Ile423-Met434	(1663)	
Gln422-Tyr435	(1657)	
Arg426-Lys432	(1681)	
Arg426-Gly431B	(1681)	
Asn425-Lys432	(1675)	
Consensus	(1681)	
	ATCAGCAGCTGCAGGCCCGCTGCTGGCCGTGGAGGGCT	
	1721	1760
Ile424-Ala433	(1709)	
Trp427-Gly431	(1721)	
Gln422-Tyr435B	(1697)	
Arg426-Gly431	(1721)	
Ile423-Met434	(1703)	
Gln422-Tyr435	(1697)	
Arg426-Lys432	(1721)	
Arg426-Gly431B	(1721)	
Asn425-Lys432	(1715)	
Consensus	(1721)	
	ACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAG	
	1761	1800
Ile424-Ala433	(1749)	
Trp427-Gly431	(1761)	
Gln422-Tyr435B	(1737)	
Arg426-Gly431	(1761)	
Ile423-Met434	(1743)	
Gln422-Tyr435	(1737)	
Arg426-Lys432	(1761)	
Arg426-Gly431B	(1761)	
Asn425-Lys432	(1755)	
Consensus	(1761)	
	CGGCAAGCTGATCTGACCCACCGCCGTGCCCTGGAACGCC	
	1801	1840
Ile424-Ala433	(1789)	
Trp427-Gly431	(1801)	
Gln422-Tyr435B	(1777)	
Arg426-Gly431	(1801)	
Ile423-Met434	(1783)	
Gln422-Tyr435	(1777)	
Arg426-Lys432	(1801)	
Arg426-Gly431B	(1801)	
Asn425-Lys432	(1795)	
Consensus	(1801)	
	AGCTGGAGCAACAAGAGCCTGGACCAGATCTGAAACAACA	
	1841	1880
Ile424-Ala433	(1829)	
Trp427-Gly431	(1841)	
Gln422-Tyr435B	(1817)	
Arg426-Gly431	(1841)	
Ile423-Met434	(1823)	
Gln422-Tyr435	(1817)	

FIG. 4I

Arg426-Lys432	(1841)	Vertebrate
Arg426-Gly431B	(1841)	Vertebrate
Asn425-Tyr432	(1835)	Vertebrate
Consensus	(1841)	TGACCTGGATGGAGTGGGAGCGCGAGATCGACAATACAC
	1881	1920
Ile424-Ala433	(1869)	Vertebrate
Trp427-Gly431	(1881)	Vertebrate
Gln422-Tyr435B	(1857)	Vertebrate
Arg426-Gly431	(1881)	Vertebrate
Ile423-Met434	(1863)	Vertebrate
Gln422-Tyr435	(1857)	Vertebrate
Arg426-Lys432	(1881)	Vertebrate
Arg426-Gly431B	(1881)	Vertebrate
Asn425-Lys432	(1875)	Vertebrate
Consensus	(1881)	CAACCTGATCTACACCTGATCGAGGAGACCCAGAACCCAG
	1921	1960
Ile424-Ala433	(1909)	Vertebrate
Trp427-Gly431	(1921)	Vertebrate
Gln422-Tyr435B	(1897)	Vertebrate
Arg426-Gly431	(1921)	Vertebrate
Ile423-Met434	(1903)	Vertebrate
Gln422-Tyr435	(1897)	Vertebrate
Arg426-Lys432	(1921)	Vertebrate
Arg426-Gly431B	(1921)	Vertebrate
Asn425-Lys432	(1915)	Vertebrate
Consensus	(1921)	CAGGAGAAGAACGAGCAGGAGCTGGAGCTGGACAAGT
	1961	2000
Ile424-Ala433	(1949)	Vertebrate
Trp427-Gly431	(1961)	Vertebrate
Gln422-Tyr435B	(1937)	Vertebrate
Arg426-Gly431	(1961)	Vertebrate
Ile423-Met434	(1943)	Vertebrate
Gln422-Tyr435	(1937)	Vertebrate
Arg426-Lys432	(1961)	Vertebrate
Arg426-Gly431B	(1961)	Vertebrate
Asn425-Lys432	(1955)	Vertebrate
Consensus	(1961)	GGGCCAGCTGTGAACTGGTTCGACATCAGCAAGTGGCT
	2001	2040
Ile424-Ala433	(1989)	Vertebrate
Trp427-Gly431	(2001)	Vertebrate
Gln422-Tyr435B	(1977)	Vertebrate
Arg426-Gly431	(2001)	Vertebrate
Ile423-Met434	(1983)	Vertebrate
Gln422-Tyr435	(1977)	Vertebrate
Arg426-Lys432	(2001)	Vertebrate
Arg426-Gly431B	(2001)	Vertebrate
Asn425-Lys432	(1995)	Vertebrate
Consensus	(2001)	GTTGGTACATCAAGATCTTCATCATGATCGTGGGGCGGCTG
	2041	2080
Ile424-Ala433	(2029)	Vertebrate
Trp427-Gly431	(2041)	Vertebrate
Gln422-Tyr435B	(2017)	Vertebrate
Arg426-Gly431	(2041)	Vertebrate
Ile423-Met434	(2023)	Vertebrate
Gln422-Tyr435	(2017)	Vertebrate
Arg426-Lys432	(2041)	Vertebrate
Arg426-Gly431B	(2041)	Vertebrate

FIG. 4J

Asn425-Lys432	(2035)	GTGGGGCTGGCCATCGTGTACCCGTCGTGACATCGTGA	2081	2120
Consensus	(2041)	GTGGGGCTGGCCATCGTGTACCCGTCGTGACATCGTGA	2081	2120
Ile424-Ala433	(2069)	ATGGCTTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2081	2120
Trp427-Gly431	(2081)	ATGGCTTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2081	2120
Gln422-Tyr435B	(2057)	ATGGCTTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2081	2120
Arg426-Gly31	(2081)	ATGGCTTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2081	2120
Ile423-Met434	(2063)	ATGGCTTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2081	2120
Gln422-Tyr435	(2057)	ATGGCTTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2081	2120
Arg426-Lys432	(2081)	ATGGCTTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2081	2120
Arg426-Gly431B	(2081)	ATGGCTTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2081	2120
Asn425-Lys432	(2075)	ATGGCTTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2081	2120
Consensus	(2081)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Ile424-Ala433	(2109)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Trp427-Gly431	(2121)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Gln422-Tyr435B	(2097)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Arg426-Gly431	(2121)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Ile423-Met434	(2103)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Gln422-Tyr435	(2097)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Arg426-Lys432	(2121)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Arg426-Gly431B	(2121)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Asn425-Lys432	(2115)	ACCGGGTGCAGGGCTACAGCCCCCTGAGCTTCCAGAC	2121	2160
Consensus	(2121)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Ile424-Ala433	(2149)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Trp427-Gly431	(2161)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Gln422-Tyr435B	(2137)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Arg426-Gly431	(2161)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Ile423-Met434	(2143)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Gln422-Tyr435	(2137)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Arg426-Lys432	(2161)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Arg426-Gly431B	(2161)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Asn425-Lys432	(2155)	CCGCTTCCCCCCCCCCCCCGGCCCCGACGGCCCCGAGGG	2161	2200
Consensus	(2161)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Ile424-Ala433	(2189)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Trp427-Gly431	(2201)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Gln422-Tyr435B	(2177)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Arg426-Gly431	(2201)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Ile423-Met434	(2183)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Gln422-Tyr435	(2177)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Arg426-Lys432	(2201)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Arg426-Gly431B	(2201)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Asn425-Lys432	(2195)	ATCGGAGGAGGGCCGGCGAGCGCGACCCGACCCGAGCA	2201	2240
Consensus	(2201)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Ile424-Ala433	(2229)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Trp427-Gly431	(2241)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Gln422-Tyr435B	(2217)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Arg426-Gly431	(2241)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Ile423-Met434	(2223)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Gln422-Tyr435	(2217)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Arg426-Lys432	(2241)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Arg426-Gly431B	(2241)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Asn425-Lys432	(2235)	GCCCCCTGGTCGACGGGCTCTGGCCCTGATCTGGGAGCA	2241	2280
Consensus	(2241)	CCTCCGGCAGGCTCTGGCCCTGATCTGGGAGCA	2280	2320

FIG. 4K

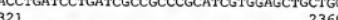
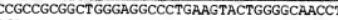
		2281	2320
Ile424-Ala433	(2269)		
Trp427-Gly431	(2281)		
Gln422-Tyr435B	(2257)		
Arg426-Gly431	(2281)		
Ile423-Met434	(2263)		
Gln422-Tyr435	(2257)		
Arg426-Lys432	(2281)		
Arg426-Gly431B	(2281)		
Asn425-Lys432	(2275)		
Consensus	(2281)	GACCTGATCTGATGCCGCCGCATGTTGAGCTGCTGG	
		2321	2360
Ile424-Ala433	(2309)		
Trp427-Gly431	(2321)		
Gln422-Tyr435B	(2297)		
Arg426-Gly431	(2321)		
Ile423-Met434	(2303)		
Gln422-Tyr435	(2297)		
Arg426-Lys432	(2321)		
Arg426-Gly431B	(2321)		
Asn425-Lys432	(2315)		
Consensus	(2321)	GCCGCCGCCGCTGGGAGGCCCTGAAGTACTGGGCAACCT	
		2361	2400
Ile424-Ala433	(2349)		
Trp427-Gly431	(2361)		
Gln422-Tyr435B	(2337)		
Arg426-Gly431	(2361)		
Ile423-Met434	(2343)		
Gln422-Tyr435	(2337)		
Arg426-Lys432	(2361)		
Arg426-Gly431B	(2361)		
Asn425-Lys432	(2355)		
Consensus	(2361)	GCTGCAGTACTGGATCCAGGAGCTGAAGAACGGCGCGTG	
		2401	2440
Ile424-Ala433	(2389)		
Trp427-Gly431	(2401)		
Gln422-Tyr435B	(2377)		
Arg426-Gly431	(2401)		
Ile423-Met434	(2383)		
Gln422-Tyr435	(2377)		
Arg426-Lys432	(2401)		
Arg426-Gly431B	(2401)		
Asn425-Lys432	(2395)		
Consensus	(2401)	AGGCTGTTGACGCCATCGCCATCGCCGTGGCGAGGGCA	
		2441	2480
Ile424-Ala433	(2429)		
Trp427-Gly431	(2441)		
Gln422-Tyr435B	(2417)		
Arg426-Gly431	(2441)		
Ile423-Met434	(2423)		
Gln422-Tyr435	(2417)		
Arg426-Lys432	(2441)		
Arg426-Gly431B	(2441)		
Asn425-Lys432	(2435)		
Consensus	(2441)	CCGACCGCATCATCGAGGTGGCCAGGCCATGGCCGCCG	
		2481	2520
Ile424-Ala433	(2469)		

FIG. 4L

Trp427-Gly431	(2481)	
Gln422-Tyr435B	(2457)	
Arg426-Gly431	(2481)	
Ile423-Met434	(2463)	
Gln422-Tyr435	(2457)	
Arg426-Lys432	(2481)	
Arg426-Gly431B	(2481)	
Asn425-Lys432	(2475)	
Consensus	(2481)	CTTCCTGCACATCCCCGGCGATCCGCCAGGGCTTCGAG -
		2521 2541
Ile424-Ala433	(2509)	
Trp427-Gly431	(2521)	
Gln422-Tyr435B	(2497)	
Arg426-Gly431	(2521)	
Ile423-Met434	(2503)	
Gln422-Tyr435	(2497)	
Arg426-Lys432	(2521)	
Arg426-Gly431B	(2521)	
Asn425-Lys432	(2515)	
Consensus	(2521)	CGCGCCCTGCTGTAACCTCGAG

FIG. 4M

	28	1	65	30
Leu122-Ser199-Tryp427-Gly431	(1)	GAATTGCCACCATGGATGCAATGAAGAGA		
Val1127-Asn195-Arg426-Gly431	(1)	GAATTGCCACCATGGATGCAATGAAGAGA		
Val1120-Thr202-Ile424-Ala433	(1)	GAATTGCCACCATGGATGCAATGAAGAGA		
Leu122-Ser199-Arg426-Lys432	(1)	GAATTGCCACCATGGATGCAATGAAGAGA		
Leu122-Ser199-Arg426-Gly431	(1)	GAATTGCCACCATGGATGCAATGAAGAGA		
Lys121-Val1200-Asn425-Lys432	(1)	GAATTGCCACCATGGATGCAATGAAGAGA		
Val1120-Ile201-Ile424-Ala433	(1)	GAATTGCCACCATGGATGCAATGAAGAGA		
Val1120-Ile201B-Ile424-Ala433	(1)	GAATTGCCACCATGGATGCAATGAAGAGA		
Consensus		31		60
Leu122-Ser199-Tryp427-Gly431	(31)	GGGCTCTGCTGTGCTGCTGTGCTGTGGA		
Val1127-Asn195-Arg426-Gly431	(31)	GGGCTCTGCTGTGCTGCTGCTGTGGA		
Val1120-Thr202-Ile424-Ala433	(31)	GGGCTCTGCTGTGCTGCTGTGGA		
Leu122-Ser199-Arg426-Lys432	(31)	GGGCTCTGCTGTGCTGCTGTGCTGTGGA		
Leu122-Ser199-Arg426-Gly431	(31)	GGGCTCTGCTGTGCTGCTGTGGA		
Lys121-Val1200-Asn425-Lys432	(31)	GGGCTCTGCTGTGCTGCTGTGCTGTGGA		
Val1120-Ile201-Ile424-Ala433	(31)	GGGCTCTGCTGTGCTGCTGTGCTGTGGA		
Val1120-Ile201B-Ile424-Ala433	(31)	GGGCTCTGCTGTGCTGCTGTGCTGTGGA		
Consensus		61		90
Leu122-Ser199-Tryp427-Gly431	(61)	GCAGCTCTCGCTTCCGCCAGGGCCGCTGGAG		
Val1127-Asn195-Arg426-Gly431	(61)	GCAGCTCTCGCTTCCGCCAGGGCCGCTGGAG		
Val1120-Thr202-Ile424-Ala433	(61)	GCAGCTCTCGCTTCCGCCAGGGCCGCTGGAG		
Leu122-Ser199-Arg426-Lys432	(61)	GCAGCTCTCGCTTCCGCCAGGGCCGCTGGAG		
Leu122-Ser199-Arg426-Gly431	(61)	GCAGCTCTCGCTTCCGCCAGGGCCGCTGGAG		
Lys121-Val1200-Asn425-Lys432	(61)	GCAGCTCTCGCTTCCGCCAGGGCCGCTGGAG		
Val1120-Ile201-Ile424-Ala433	(61)	GCAGCTCTCGCTTCCGCCAGGGCCGCTGGAG		
Val1120-Ile201B-Ile424-Ala433	(61)	GCAGCTCTCGCTTCCGCCAGGGCCGCTGGAG		
Consensus		91		120
Leu122-Ser199-Tryp427-Gly431	(91)	AAGCTGTGGTGACCGTGTACTACGGGGTG		
Val1127-Asn195-Arg426-Gly431	(91)	AAGCTGTGGTGACCGTGTACTACGGGGTG		
Val1120-Thr202-Ile424-Ala433	(91)	AAGCTGTGGTGACCGTGTACTACGGGGTG		
Leu122-Ser199-Arg426-Lys432	(91)	AAGCTGTGGTGACCGTGTACTACGGGGTG		
Leu122-Ser199-Arg426-Gly431	(91)	AAGCTGTGGTGACCGTGTACTACGGGGTG		
Lys121-Val1200-Asn425-Lys432	(91)	AAGCTGTGGTGACCGTGTACTACGGGGTG		
Val1120-Ile201-Ile424-Ala433	(91)	AAGCTGTGGTGACCGTGTACTACGGGGTG		
Val1120-Ile201B-Ile424-Ala433	(91)	AAGCTGTGGTGACCGTGTACTACGGGGTG		
Consensus		121		150
Leu122-Ser199-Tryp427-Gly431	(121)	CCCGTGTGGAAAGGGGCCACACACCGCTG		
Val1127-Asn195-Arg426-Gly431	(121)	CCCGTGTGGAAAGGGGCCACACACCGCTG		
Val1120-Thr202-Ile424-Ala433	(121)	CCCGTGTGGAAAGGGGCCACACACCGCTG		
Leu122-Ser199-Arg426-Lys432	(121)	CCCGTGTGGAAAGGGGCCACACACCGCTG		
Leu122-Ser199-Arg426-Gly431	(121)	CCCGTGTGGAAAGGGGCCACACACCGCTG		
Lys121-Val1200-Asn425-Lys432	(121)	CCCGTGTGGAAAGGGGCCACACACCGCTG		
Val1120-Ile201-Ile424-Ala433	(121)	CCCGTGTGGAAAGGGGCCACACACCGCTG		
Val1120-Ile201B-Ile424-Ala433	(121)	CCCGTGTGGAAAGGGGCCACACACCGCTG		
Consensus		151		180
Leu122-Ser199-Tryp427-Gly431	(151)	TTCCTGCCACGGGACGCCAAGGCCCTACGAC		
Val1127-Asn195-Arg426-Gly431	(151)	TTCCTGCCACGGGACGCCAAGGCCCTACGAC		
Val1120-Thr202-Ile424-Ala433	(151)	TTCCTGCCACGGGACGCCAAGGCCCTACGAC		
Leu122-Ser199-Arg426-Lys432	(151)	TTCCTGCCACGGGACGCCAAGGCCCTACGAC		
Leu122-Ser199-Arg426-Gly431	(151)	TTCCTGCCACGGGACGCCAAGGCCCTACGAC		
Lys121-Val1200-Asn425-Lys432	(151)	TTCCTGCCACGGGACGCCAAGGCCCTACGAC		

FIG. 5A

Val120-Ile201-Ile424-Ala433	(151)	TTCCTGCGCCAGGGCGCCAAAGGCCCTACGAG
Val120-Ile201B-Ile424-Ala433	(151)	TTCCTGCGCCAGGGCGCCAAAGGCCCTACGAG
Consensus	(151)	TTCCTGCGCCAGGGCGCCAAAGGCCCTACGAG
	181	181
Leu122-Ser199-Tryp427-Gly431	(181)	ACCGAGGTGCAACACGTGTGGGCCACCCAG
Val122-Asn195-Arg426-Gly431	(181)	ACCGAGGTGCAACACGTGTGGGCCACCCAG
Val120-Thr202-Ile424-Ala433	(181)	ACCGAGGTGCAACACGTGTGGGCCACCCAG
Leu122-Ser199-Arg426-Lys432	(181)	ACCGAGGTGCAACACGTGTGGGCCACCCAG
Leu122-Ser199-Arg426-Gly431	(181)	ACCGAGGTGCAACACGTGTGGGCCACCCAG
Lys121-Val200-Asn425-Lys432	(181)	ACCGAGGTGCAACACGTGTGGGCCACCCAG
Val120-Ile201-Ile424-Ala433	(181)	ACCGAGGTGCAACACGTGTGGGCCACCCAG
Val120-Ile201B-Ile424-Ala433	(181)	ACCGAGGTGCAACACGTGTGGGCCACCCAG
Consensus	(181)	ACCGAGGTGCAACACGTGTGGGCCACCCAG
	211	240
Leu122-Ser199-Tryp427-Gly431	(211)	GCCCTGGTGGCCACCGACCCCCAACCCCCAG
Val122-Asn195-Arg426-Gly431	(211)	GCCCTGGTGGCCACCGACCCCCAACCCCCAG
Val120-Thr202-Ile424-Ala433	(211)	GCCCTGGTGGCCACCGACCCCCAACCCCCAG
Leu122-Ser199-Arg426-Lys432	(211)	GCCCTGGTGGCCACCGACCCCCAACCCCCAG
Leu122-Ser199-Arg426-Gly431	(211)	GCCCTGGTGGCCACCGACCCCCAACCCCCAG
Lys121-Val200-Asn425-Lys432	(211)	GCCCTGGTGGCCACCGACCCCCAACCCCCAG
Val120-Ile201-Ile424-Ala433	(211)	GCCCTGGTGGCCACCGACCCCCAACCCCCAG
Val120-Ile201B-Ile424-Ala433	(211)	GCCCTGGTGGCCACCGACCCCCAACCCCCAG
Consensus	(211)	GCCCTGGTGGCCACCGACCCCCAACCCCCAG
	241	270
Leu122-Ser199-Tryp427-Gly431	(241)	GAGATCTGCTGGAGACCTGACCGAGAAC
Val122-Asn195-Arg426-Gly431	(241)	GAGATCTGCTGGAGACCTGACCGAGAAC
Val120-Thr202-Ile424-Ala433	(241)	GAGATCTGCTGGAGACCTGACCGAGAAC
Leu122-Ser199-Arg426-Lys432	(241)	GAGATCTGCTGGAGACCTGACCGAGAAC
Leu122-Ser199-Arg426-Gly431	(241)	GAGATCTGCTGGAGACCTGACCGAGAAC
Lys121-Val200-Asn425-Lys432	(241)	GAGATCTGCTGGAGACCTGACCGAGAAC
Val120-Ile201-Ile424-Ala433	(241)	GAGATCTGCTGGAGACCTGACCGAGAAC
Val120-Ile201B-Ile424-Ala433	(241)	GAGATCTGCTGGAGACCTGACCGAGAAC
Consensus	(241)	GAGATCTGCTGGAGACCTGACCGAGAAC
	271	300
Leu122-Ser199-Tryp427-Gly431	(271)	TTCACATGTTGGAAAGACAAACATGGTGAG
Val122-Asn195-Arg426-Gly431	(271)	TTCACATGTTGGAAAGACAAACATGGTGAG
Val120-Thr202-Ile424-Ala433	(271)	TTCACATGTTGGAAAGACAAACATGGTGAG
Leu122-Ser199-Arg426-Lys432	(271)	TTCACATGTTGGAAAGACAAACATGGTGAG
Leu122-Ser199-Arg426-Gly431	(271)	TTCACATGTTGGAAAGACAAACATGGTGAG
Lys121-Val200-Asn425-Lys432	(271)	TTCACATGTTGGAAAGACAAACATGGTGAG
Val120-Ile201-Ile424-Ala433	(271)	TTCACATGTTGGAAAGACAAACATGGTGAG
Val120-Ile201B-Ile424-Ala433	(271)	TTCACATGTTGGAAAGACAAACATGGTGAG
Consensus	(271)	TTCACATGTTGGAAAGACAAACATGGTGAG
	301	330
Leu122-Ser199-Tryp427-Gly431	(301)	CAGATGCCAGGGACATCTCATGCCCTGG
Val122-Asn195-Arg426-Gly431	(301)	CAGATGCCAGGGACATCTCATGCCCTGG
Val120-Thr202-Ile424-Ala433	(301)	CAGATGCCAGGGACATCTCATGCCCTGG
Leu122-Ser199-Arg426-Lys432	(301)	CAGATGCCAGGGACATCTCATGCCCTGG
Leu122-Ser199-Arg426-Gly431	(301)	CAGATGCCAGGGACATCTCATGCCCTGG
Lys121-Val200-Asn425-Lys432	(301)	CAGATGCCAGGGACATCTCATGCCCTGG
Val120-Ile201-Ile424-Ala433	(301)	CAGATGCCAGGGACATCTCATGCCCTGG
Val120-Ile201B-Ile424-Ala433	(301)	CAGATGCCAGGGACATCTCATGCCCTGG
Consensus	(301)	CAGATGCCAGGGACATCTCATGCCCTGG
	331	360
Leu122-Ser199-Tryp427-Gly431	(331)	GACCGAGGCTGAGGCCCTCGCTGAGCTG
Val122-Asn195-Arg426-Gly431	(331)	GACCGAGGCTGAGGCCCTCGCTGAGCTG
Val120-Thr202-Ile424-Ala433	(331)	GACCGAGGCTGAGGCCCTCGCTGAGCTG

FIG. 5R

Leu122-Ser199-Arg426-Lys432	(331) GACCAGAGCCTGAAGCCTCGCTGAAGCTG
Leu122-Ser199-Arg426-Gly431	(331) GACCAGAGCCTGAAGCCTCGCTGAAGCTG
Lys121-Val200-Asn425-Lys432	(331) GACCAGAGCCTGAAGCCTCGCTGAAGCTG
Val120-Ile201-Ile424-Ala433	(331) GACCAGAGCCTGAAGCCTCGCTGAAGCTG
Val120-Ile201B-Ile424-Ala433	(331) GACCAGAGCCTGAAGCCTCGCTGAAGCTG
Consensus	361 390
Leu122-Ser199-Tryp427-Gly431	(361) -----GG-----
Val127-Asn195-Arg426-Gly431	(361) ACCCCCCCTGTGCGTGGGGCAGGGAACTGC
Val120-Thr202-Ile424-Ala433	(355) -----GG-----
Leu122-Ser199-Arg426-Lys432	(361) -----GG-----
Leu122-Ser199-Arg426-Gly431	(361) -----GG-----
Lys121-Val200-Asn425-Lys432	(357) -----GG-----
Val120-Ile201-Ile424-Ala433	(355) -----GG-----
Val120-Ile201B-Ile424-Ala433	(355) -----GG-----
Consensus	(361) GG 420
Leu122-Ser199-Tryp427-Gly431	(363) --AACAGGGTGTATCACCCAGGCTGCCCC
Val127-Asn195-Arg426-Gly431	(391) AACACCAAGGGTGTATCACCCAGGCTGCCCC
Val120-Thr202-Ile424-Ala433	(357) ---CGGGC---CACCCAGGCTGCCCC
Leu122-Ser199-Arg426-Lys432	(363) --CAACAGGGTGTATCACCCAGGCTGCCCC
Leu122-Ser199-Arg426-Gly431	(363) --CAAAAGGGTGTATCACCCAGGCTGCCCC
Lys121-Val200-Asn425-Lys432	(359) ---CCCGGGTGTATCACCCAGGCTGCCCC
Val120-Ile201-Ile424-Ala433	(355) -----GGGGCATCACCCAGGCTGCCCC
Val120-Ile201B-Ile424-Ala433	(355) -----CCGGGCATCACCCAGGCTGCCCC
Consensus	(391) CA CAGGGTGTATCACCCAGGCTGCCCC 450
Leu122-Ser199-Tryp427-Gly431	(391) AAGGTGAGCTTCGAGCCCCATCCCCATCCCC
Val127-Asn195-Arg426-Gly431	(421) AAGGTGAGCTTCGAGCCCCATCCCCATCCCC
Val120-Thr202-Ile424-Ala433	(379) AAGGTGAGCTTCGAGCCCCATCCCCATCCCC
Leu122-Ser199-Arg426-Lys432	(391) AAGGTGAGCTTCGAGCCCCATCCCCATCCCC
Leu122-Ser199-Arg426-Gly431	(391) AAGGTGAGCTTCGAGCCCCATCCCCATCCCC
Lys121-Val200-Asn425-Lys432	(385) AAGGTGAGCTTCGAGCCCCATCCCCATCCCC
Val120-Ile201-Ile424-Ala433	(379) AAGGTGAGCTTCGAGCCCCATCCCCATCCCC
Val120-Ile201B-Ile424-Ala433	(379) AAGGTGAGCTTCGAGCCCCATCCCCATCCCC
Consensus	(421) AAGGTGAGCTTCGAGCCCCATCCCCATCCCC 480
Leu122-Ser199-Tryp427-Gly431	(421) TACTGGCGGGCGGGGCTTCGGCATCTCTG
Val127-Asn195-Arg426-Gly431	(451) TACTGGCGGGCGGGGCTTCGGCATCTCTG
Val120-Thr202-Ile424-Ala433	(409) TACTGGCGGGCGGGGCTTCGGCATCTCTG
Leu122-Ser199-Arg426-Lys432	(421) TACTGGCGGGCGGGGCTTCGGCATCTCTG
Leu122-Ser199-Arg426-Gly431	(421) TACTGGCGGGCGGGGCTTCGGCATCTCTG
Lys121-Val200-Asn425-Lys432	(415) TACTGGCGGGCGGGGCTTCGGCATCTCTG
Val120-Ile201-Ile424-Ala433	(409) TACTGGCGGGCGGGGCTTCGGCATCTCTG
Val120-Ile201B-Ile424-Ala433	(409) TACTGGCGGGCGGGGCTTCGGCATCTCTG
Consensus	(451) TACTGGCGGGCGGGGCTTCGGCATCTCTG 481 510
Leu122-Ser199-Tryp427-Gly431	(451) AACTGCAACGACAAGAAGTTCAACGGCAGC
Val127-Asn195-Arg426-Gly431	(481) AAGTGCACGACAAGAAGTTCAACGGCAGC
Val120-Thr202-Ile424-Ala433	(439) AAGTGCACGACAAGAAGTTCAACGGCAGC
Leu122-Ser199-Arg426-Lys432	(451) AAGTGCACGACAAGAAGTTCAACGGCAGC
Leu122-Ser199-Arg426-Gly431	(451) AAGTGCACGACAAGAAGTTCAACGGCAGC
Lys121-Val200-Asn425-Lys432	(445) AAGTGCACGACAAGAAGTTCAACGGCAGC
Val120-Ile201-Ile424-Ala433	(439) AAGTGCACGACAAGAAGTTCAACGGCAGC
Val120-Ile201B-Ile424-Ala433	(439) AAGTGCACGACAAGAAGTTCAACGGCAGC
Consensus	(481) AAGTGCACGACAAGAAGTTCAACGGCAGC 511 540

FIG. 5C

Leu122-Ser199-Tryp427-Gly431	(481)	GGCCCCCTGCACCAACGTGAGCACCGTGAG
Val1127-Asn195-Arg426-Gly431	(511)	GGCCCCCTGCACCAACGTGAGCACCGTGAG
Val1120-Thr202-Ile424-Ala433	(469)	GGCCCCCTGCACCAACGTGAGCACCGTGAG
Leu122-Ser199-Arg426-Lys432	(481)	GGCCCCCTGCACCAACGTGAGCACCGTGAG
Leu122-Ser199-Arg426-Gly431	(481)	GGCCCCCTGCACCAACGTGAGCACCGTGAG
Lys121-Val200-Asn425-Lys432	(475)	GGCCCCCTGCACCAACGTGAGCACCGTGAG
Val1120-Ile201-Ile424-Ala433	(469)	GGCCCCCTGCACCAACGTGAGCACCGTGAG
Val1120-Ile201B-Ile424-Ala433	(469)	GGCCCCCTGCACCAACGTGAGCACCGTGAG
Consensus	(511)	GGCCCCCTGCACCAACGTGAGCACCGTGAG
	541	570
Leu122-Ser199-Tryp427-Gly431	(511)	TGCACCCACGGCATCCGCCCGTGGTGAGC
Val1127-Asn195-Arg426-Gly431	(541)	TGCACCCACGGCATCCGCCCGTGGTGAGC
Val1120-Thr202-Ile424-Ala433	(499)	TGCACCCACGGCATCCGCCCGTGGTGAGC
Leu122-Ser199-Arg426-Lys432	(511)	TGCACCCACGGCATCCGCCCGTGGTGAGC
Leu122-Ser199-Arg426-Gly431	(511)	TGCACCCACGGCATCCGCCCGTGGTGAGC
Lys121-Val200-Asn425-Lys432	(505)	TGCACCCACGGCATCCGCCCGTGGTGAGC
Val1120-Ile201-Ile424-Ala433	(499)	TGCACCCACGGCATCCGCCCGTGGTGAGC
Val1120-Ile201B-Ile424-Ala433	(499)	TGCACCCACGGCATCCGCCCGTGGTGAGC
Consensus	(541)	TGCACCCACGGCATCCGCCCGTGGTGAGC
	571	600
Leu122-Ser199-Tryp427-Gly431	(541)	ACCAGCTGCTGCTGAACGGCAGCCCTGGCC
Val1127-Asn195-Arg426-Gly431	(571)	ACCAGCTGCTGCTGAACGGCAGCCCTGGCC
Val1120-Thr202-Ile424-Ala433	(529)	ACCAGCTGCTGCTGAACGGCAGCCCTGGCC
Leu122-Ser199-Arg426-Lys432	(541)	ACCAGCTGCTGCTGAACGGCAGCCCTGGCC
Leu122-Ser199-Arg426-Gly431	(541)	ACCAGCTGCTGCTGAACGGCAGCCCTGGCC
Lys121-Val200-Asn425-Lys432	(535)	ACCAGCTGCTGCTGAACGGCAGCCCTGGCC
Val1120-Ile201-Ile424-Ala433	(529)	ACCAGCTGCTGCTGAACGGCAGCCCTGGCC
Val1120-Ile201B-Ile424-Ala433	(529)	ACCAGCTGCTGCTGAACGGCAGCCCTGGCC
Consensus	(571)	ACCAGCTGCTGCTGAACGGCAGCCCTGGCC
	601	630
Leu122-Ser199-Tryp427-Gly431	(571)	GAGGAGGGCGTGGTGATCCGCAGCGAGAAC
Val1127-Asn195-Arg426-Gly431	(601)	GAGGAGGGCGTGGTGATCCGCAGCGAGAAC
Val1120-Thr202-Ile424-Ala433	(559)	GAGGAGGGCGTGGTGATCCGCAGCGAGAAC
Leu122-Ser199-Arg426-Lys432	(571)	GAGGAGGGCGTGGTGATCCGCAGCGAGAAC
Leu122-Ser199-Arg426-Gly431	(571)	GAGGAGGGCGTGGTGATCCGCAGCGAGAAC
Lys121-Val200-Asn425-Lys432	(565)	GAGGAGGGCGTGGTGATCCGCAGCGAGAAC
Val1120-Ile201-Ile424-Ala433	(559)	GAGGAGGGCGTGGTGATCCGCAGCGAGAAC
Val1120-Ile201B-Ile424-Ala433	(559)	GAGGAGGGCGTGGTGATCCGCAGCGAGAAC
Consensus	(601)	GAGGAGGGCGTGGTGATCCGCAGCGAGAAC
	631	660
Leu122-Ser199-Tryp427-Gly431	(601)	TTCRCCGACAACGCCAACGACCATCATCGTG
Val1127-Asn195-Arg426-Gly431	(631)	TTCRCCGACAACGCCAACGACCATCATCGTG
Val1120-Thr202-Ile424-Ala433	(589)	TTCRCCGACAACGCCAACGACCATCATCGTG
Leu122-Ser199-Arg426-Lys432	(601)	TTCRCCGACAACGCCAACGACCATCATCGTG
Leu122-Ser199-Arg426-Gly431	(601)	TTCRCCGACAACGCCAACGACCATCATCGTG
Lys121-Val200-Asn425-Lys432	(595)	TTCRCCGACAACGCCAACGACCATCATCGTG
Val1120-Ile201-Ile424-Ala433	(589)	TTCRCCGACAACGCCAACGACCATCATCGTG
Val1120-Ile201B-Ile424-Ala433	(589)	TTCRCCGACAACGCCAACGACCATCATCGTG
Consensus	(631)	TTCRCCGACAACGCCAACGACCATCATCGTG
	661	690
Leu122-Ser199-Tryp427-Gly431	(631)	CAGCTGAGGGAGAGGGTGGAGATCACTGC
Val1127-Asn195-Arg426-Gly431	(661)	CAGCTGAGGGAGAGGGTGGAGATCACTGC
Val1120-Thr202-Ile424-Ala433	(619)	CAGCTGAGGGAGAGGGTGGAGATCACTGC
Leu122-Ser199-Arg426-Lys432	(631)	CAGCTGAGGGAGAGGGTGGAGATCACTGC
Leu122-Ser199-Arg426-Gly431	(631)	CAGCTGAGGGAGAGGGTGGAGATCACTGC
Lys121-Val200-Asn425-Lys432	(625)	CAGCTGAGGGAGAGGGTGGAGATCACTGC
Val1120-Ile201-Ile424-Ala433	(619)	CAGCTGAGGGAGAGGGTGGAGATCACTGC

Val1120-Ile201-Ile424-Ala433	613) CAGCTGAAGGAGAGCTGGAGATCACTGCG
Consensus	(661) CAGCTGAAGGAGAGCTGGAGATCACTGCG
	691 720
Leu122-Ser199-Trypt427-Gly431	(661) ACCGGCCCCAACAAACACACCCGCAAGAGC
Val127-Asn195-Arg426-Gly431	(691) ACCGGCCCCAACAAACACACCCGCAAGAGC
Val120-Thr202-Ile424-Ala433	(649) ACCGGCCCCAACAAACACACCCGCAAGAGC
Leu122-Ser199-Arg426-Lys432	(661) ACCGGCCCCAACAAACACACCCGCAAGAGC
Leu122-Ser199-Arg426-Gly431	(661) ACCGGCCCCAACAAACACACCCGCAAGAGC
Lys121-Val200-Asn425-Lys432	(655) ACCGGCCCCAACAAACACACCCGCAAGAGC
Val120-Ile201-Ile424-Ala433	(649) ACCGGCCCCAACAAACACACCCGCAAGAGC
Val120-Ile201B-Ile424-Ala433	(649) ACCGGCCCCAACAAACACACCCGCAAGAGC
Consensus	(691) ACCGGCCCCAACAAACACACCCGCAAGAGC
	721 750
Leu122-Ser199-Trypt427-Gly431	(691) ATCACCATCGGCCCCGGCGCGCTTCTAC
Val127-Asn195-Arg426-Gly431	(721) ATCACCATCGGCCCCGGCGCGCTTCTAC
Val120-Thr202-Ile424-Ala433	(679) ATCACCATCGGCCCCGGCGCGCTTCTAC
Leu122-Ser199-Arg426-Lys432	(691) ATCACCATCGGCCCCGGCGCGCTTCTAC
Leu122-Ser199-Arg426-Gly431	(691) ATCACCATCGGCCCCGGCGCGCTTCTAC
Lys121-Val200-Asn425-Lys432	(685) ATCACCATCGGCCCCGGCGCGCTTCTAC
Val120-Ile201-Ile424-Ala433	(679) ATCACCATCGGCCCCGGCGCGCTTCTAC
Val120-Ile201B-Ile424-Ala433	(679) ATCACCATCGGCCCCGGCGCGCTTCTAC
Consensus	(721) ATCACCATCGGCCCCGGCGCGCTTCTAC
	751 780
Leu122-Ser199-Trypt427-Gly431	(721) GCAACGGCGGACATCATCGGCGACATCGGC
Val127-Asn195-Arg426-Gly431	(751) GCAACGGCGGACATCATCGGCGACATCGGC
Val120-Thr202-Ile424-Ala433	(709) GCAACGGCGGACATCATCGGCGACATCGGC
Leu122-Ser199-Arg426-Lys432	(721) GCAACGGCGGACATCATCGGCGACATCGGC
Leu122-Ser199-Arg426-Gly431	(721) GCAACGGCGGACATCATCGGCGACATCGGC
Lys121-Val200-Asn425-Lys432	(715) GCAACGGCGGACATCATCGGCGACATCGGC
Val120-Ile201-Ile424-Ala433	(709) GCAACGGCGGACATCATCGGCGACATCGGC
Val120-Ile201B-Ile424-Ala433	(709) GCAACGGCGGACATCATCGGCGACATCGGC
Consensus	(751) GCAACGGCGGACATCATCGGCGACATCGGC
	781 810
Leu122-Ser199-Trypt427-Gly431	(751) CAGGGCCACTGCACACATCAGCGGCGAGAAG
Val127-Asn195-Arg426-Gly431	(781) CAGGGCCACTGCACACATCAGCGGCGAGAAG
Val120-Thr202-Ile424-Ala433	(739) CAGGGCCACTGCACACATCAGCGGCGAGAAG
Leu122-Ser199-Arg426-Lys432	(751) CAGGGCCACTGCACACATCAGCGGCGAGAAG
Leu122-Ser199-Arg426-Gly431	(751) CAGGGCCACTGCACACATCAGCGGCGAGAAG
Lys121-Val200-Asn425-Lys432	(745) CAGGGCCACTGCACACATCAGCGGCGAGAAG
Val120-Ile201-Ile424-Ala433	(739) CAGGGCCACTGCACACATCAGCGGCGAGAAG
Val120-Ile201B-Ile424-Ala433	(739) CAGGGCCACTGCACACATCAGCGGCGAGAAG
Consensus	(781) CAGGGCCACTGCACACATCAGCGGCGAGAAG
	811 840
Leu122-Ser199-Trypt427-Gly431	(781) TGGACACACCCCTGAAGCAGATCTGTGAC
Val127-Asn195-Arg426-Gly431	(811) TGGACACACCCCTGAAGCAGATCTGTGAC
Val120-Thr202-Ile424-Ala433	(769) TGGACACACCCCTGAAGCAGATCTGTGAC
Leu122-Ser199-Arg426-Lys432	(781) TGGACACACCCCTGAAGCAGATCTGTGAC
Leu122-Ser199-Arg426-Gly431	(781) TGGACACACCCCTGAAGCAGATCTGTGAC
Lys121-Val200-Asn425-Lys432	(775) TGGACACACCCCTGAAGCAGATCTGTGAC
Val120-Ile201-Ile424-Ala433	(769) TGGACACACCCCTGAAGCAGATCTGTGAC
Val120-Ile201B-Ile424-Ala433	(769) TGGACACACCCCTGAAGCAGATCTGTGAC
Consensus	(811) TGGACACACCCCTGAAGCAGATCTGTGAC
	841 870
Leu122-Ser199-Trypt427-Gly431	(811) AAGCTGCAGGCCAGTTCTGGCAACAGAGC
Val127-Asn195-Arg426-Gly431	(841) AAGCTGCAGGCCAGTTCTGGCAACAGAGC
Val120-Thr202-Ile424-Ala433	(799) AAGCTGCAGGCCAGTTCTGGCAACAGAGC
Leu122-Ser199-Arg426-Lys432	(811) AAGCTGCAGGCCAGTTCTGGCAACAGAGC

FIG. 5E

	33	/	65
Leu122-Ser199-Arg426-Gly431	(811)	AAGCTGCAGGCCCACTTGGCAACAAGACC	
Lys121-Val200-Asn425-Lys432	(805)	AAGCTGCAGGCCCACTTGGCAACAAGACC	
Val120-Ile201-Ile424-Ala433	(799)	AAGCTGCAGGCCCACTTGGCAACAAGACC	
Val120-Ile201B-Ile424-Ala433	(799)	AAGCTGCAGGCCCACTTGGCAACAAGACC	
Consensus	(811)	AAGCTGCAGGCCCACTTGGCAACAAGACC	
	871		900
Leu122-Ser199-Tryp427-Gly431	(841)	ATCGTGTCAAGCAGAGCACGGCGGCGAC	
Val127-Asn195-Arg426-Gly431	(871)	ATCGTGTCAAGCAGAGCACGGCGGCGAC	
Val120-Thr202-Ile424-Ala433	(829)	ATCGTGTCAAGCAGAGCACGGCGGCGAC	
Leu122-Ser199-Arg426-Lys432	(841)	ATCGTGTCAAGCAGAGCACGGCGGCGAC	
Leu122-Ser199-Arg426-Gly431	(841)	ATCGTGTCAAGCAGAGCACGGCGGCGAC	
Lys121-Val200-Asn425-Lys432	(835)	ATCGTGTCAAGCAGAGCACGGCGGCGAC	
Val120-Ile201-Ile424-Ala433	(829)	ATCGTGTCAAGCAGAGCACGGCGGCGAC	
Val120-Ile201B-Ile424-Ala433	(829)	ATCGTGTCAAGCAGAGCACGGCGGCGAC	
Consensus	(871)	ATCGTGTCAAGCAGAGCACGGCGGCGAC	
	901		930
Leu122-Ser199-Tryp427-Gly431	(871)	CCGGAGATCGTGTATGCACAGCTTCACTGC	
Val127-Asn195-Arg426-Gly431	(901)	CCGGAGATCGTGTATGCACAGCTTCACTGC	
Val120-Thr202-Ile424-Ala433	(859)	CCGGAGATCGTGTATGCACAGCTTCACTGC	
Leu122-Ser199-Arg426-Lys432	(871)	CCGGAGATCGTGTATGCACAGCTTCACTGC	
Leu122-Ser199-Arg426-Gly431	(871)	CCGGAGATCGTGTATGCACAGCTTCACTGC	
Lys121-Val200-Asn425-Lys432	(865)	CCGGAGATCGTGTATGCACAGCTTCACTGC	
Val120-Ile201-Ile424-Ala433	(859)	CCGGAGATCGTGTATGCACAGCTTCACTGC	
Val120-Ile201B-Ile424-Ala433	(859)	CCGGAGATCGTGTATGCACAGCTTCACTGC	
Consensus	(901)	CCGGAGATCGTGTATGCACAGCTTCACTGC	
	931		960
Leu122-Ser199-Tryp427-Gly431	(901)	GGGGGGGAGTTCTCTACTGCACAGCACCC	
Val127-Asn195-Arg426-Gly431	(931)	GGGGGGGAGTTCTCTACTGCACAGCACCC	
Val120-Thr202-Ile424-Ala433	(889)	GGGGGGGAGTTCTCTACTGCACAGCACCC	
Leu122-Ser199-Arg426-Lys432	(901)	GGGGGGGAGTTCTCTACTGCACAGCACCC	
Leu122-Ser199-Arg426-Gly431	(901)	GGGGGGGAGTTCTCTACTGCACAGCACCC	
Lys121-Val200-Asn425-Lys432	(895)	GGGGGGGAGTTCTCTACTGCACAGCACCC	
Val120-Ile201-Ile424-Ala433	(889)	GGGGGGGAGTTCTCTACTGCACAGCACCC	
Val120-Ile201B-Ile424-Ala433	(889)	GGGGGGGAGTTCTCTACTGCACAGCACCC	
Consensus	(931)	GGGGGGGAGTTCTCTACTGCACAGCACCC	
	961		990
Leu122-Ser199-Tryp427-Gly431	(931)	CAGCTGTTCAACACGCCCTGGAACACACCC	
Val127-Asn195-Arg426-Gly431	(961)	CAGCTGTTCAACACGCCCTGGAACACACCC	
Val120-Thr202-Ile424-Ala433	(919)	CAGCTGTTCAACACGCCCTGGAACACACCC	
Leu122-Ser199-Arg426-Lys432	(931)	CAGCTGTTCAACACGCCCTGGAACACACCC	
Leu122-Ser199-Arg426-Gly431	(931)	CAGCTGTTCAACACGCCCTGGAACACACCC	
Lys121-Val200-Asn425-Lys432	(925)	CAGCTGTTCAACACGCCCTGGAACACACCC	
Val120-Ile201-Ile424-Ala433	(919)	CAGCTGTTCAACACGCCCTGGAACACACCC	
Val120-Ile201B-Ile424-Ala433	(919)	CAGCTGTTCAACACGCCCTGGAACACACCC	
Consensus	(961)	CAGCTGTTCAACACGCCCTGGAACACACCC	
	991		1020
Leu122-Ser199-Tryp427-Gly431	(961)	ATCGGCCCCAACACACCCAACGGCACCTTC	
Val127-Asn195-Arg426-Gly431	(991)	ATCGGCCCCAACACACCCAACGGCACCTTC	
Val120-Thr202-Ile424-Ala433	(949)	ATCGGCCCCAACACACCCAACGGCACCTTC	
Leu122-Ser199-Arg426-Lys432	(961)	ATCGGCCCCAACACACCCAACGGCACCTTC	
Leu122-Ser199-Arg426-Gly431	(961)	ATCGGCCCCAACACACCCAACGGCACCTTC	
Lys121-Val200-Asn425-Lys432	(955)	ATCGGCCCCAACACACCCAACGGCACCTTC	
Val120-Ile201-Ile424-Ala433	(949)	ATCGGCCCCAACACACCCAACGGCACCTTC	
Val120-Ile201B-Ile424-Ala433	(949)	ATCGGCCCCAACACACCCAACGGCACCTTC	
Consensus	(991)	ATCGGCCCCAACACACCCAACGGCACCTTC	
	1021		1050
Leu122-Ser199-Tryp427-Gly431	(991)	ACCCCTGCCCCATGGCCATCAAGCAGATCATC	

FIG. 5F

Val1127-Asn195-Arg426-Gly431	(1021) ACCCCTGCCCTGCCGCATCAAGCAGATCATC
Val1120-Thr202-Ile424-Ala433	(1079) ACCCCTGCCCTGCCGCATCAAGCAGATCATC
Leu122-Ser199-Arg426-Lys432	(991) ACCCCTGCCCTGCCGCATCAAGCAGATCATC
Leu122-Ser199-Arg426-Gly431	(991) ACCCCTGCCCTGCCGCATCAAGCAGATCATC
Lys121-Val200-Asn425-Lys432	(985) ACCCCTGCCCTGCCGCATCAAGCAGATCATC
Val1120-Ile201-Ile424-Ala433	(979) ACCCCTGCCCTGCCGCATCAAGCAGATCATC
Val1120-Ile201B-Ile424-Ala433	(979) ACCCCTGCCCTGCCGCATCAAGCAGATCATC
Consensus	(1021) ACCCCTGCCCTGCCGCATCAAGCAGATCATC
	1080
Leu122-Ser199 Tryp427-Gly431	(1021) ACCCGCTGGGGGGCAAGGCCATGTACGCC
Val1127-Asn195-Arg426-Gly431	(1051) ACCCGCGGGGGGGCAAGGCCATGTACGCC
Val1120-Thr202-Ile424-Ala433	(1009) -----GGGGC-----GCCATGTACGCC
Leu122-Ser199-Arg426-Lys432	(1021) ACCCGCGGGGGCAAGGCCATGTACGCC
Leu122-Ser199-Arg426-Gly431	(1021) ACCCGCGGCAGGGCAAGGCCATGTACGCC
Lys121-Val200-Asn425-Lys432	(1015) AC-----GCCCAAGGCCATGTACGCC
Val1120-Ile201-Ile424-Ala433	(1009) -----GGGGC-----GCCATGTACGCC
Val1120-Ile201B-Ile424-Ala433	(1009) -----GGGGC-----GCCATGTACGCC
Consensus	(1051) ACCCGGG G GGGGGCAAGGCCATGTACGCC
	1110
Leu122-Ser199 Tryp427-Gly431	(1051) CCCCCATCGGGGGCAGATCCGCTCGAGC
Val1127-Asn195-Arg426-Gly431	(1081) CCCCCATCGGGGGCAGATCCGCTCGAGC
Val1120-Thr202-Ile424-Ala433	(1027) CCCCCATCGGGGGCAGATCCGCTCGAGC
Leu122-Ser199-Arg426-Lys432	(1051) CCCCCATCGGGGGCAGATCCGCTCGAGC
Leu122-Ser199-Arg426-Gly431	(1051) CCCCCATCGGGGGCAGATCCGCTCGAGC
Lys121-Val200-Asn425-Lys432	(1039) CCCCCATCGGGGGCAGATCCGCTCGAGC
Val1120-Ile201-Ile424-Ala433	(1027) CCCCCATCGGGGGCAGATCCGCTCGAGC
Val1120-Ile201B-Ile424-Ala433	(1027) CCCCCATCGGGGGCAGATCCGCTCGAGC
Consensus	(1081) CCCCCATCGGGGGCAGATCCGCTCGAGC
	1140
Leu122-Ser199 Tryp427-Gly431	(1081) ACCAACATCACGGGCCCTGCTGACCCCC
Val1127-Asn195-Arg426-Gly431	(1111) ACCAACATCACGGGCCCTGCTGACCCCC
Val1120-Thr202-Ile424-Ala433	(1057) ACCAACATCACGGGCCCTGCTGACCCCC
Leu122-Ser199-Arg426-Lys432	(1081) ACCAACATCACGGGCCCTGCTGACCCCC
Leu122-Ser199-Arg426-Gly431	(1081) ACCAACATCACGGGCCCTGCTGACCCCC
Lys121-Val200-Asn425-Lys432	(1069) ACCAACATCACGGGCCCTGCTGACCCCC
Val1120-Ile201-Ile424-Ala433	(1057) ACCAACATCACGGGCCCTGCTGACCCCC
Val1120-Ile201B-Ile424-Ala433	(1057) ACCAACATCACGGGCCCTGCTGACCCCC
Consensus	(1111) ACCAACATCACGGGCCCTGCTGACCCCC
	1170
Leu122-Ser199 Tryp427-Gly431	(1111) GACGGGGCAAGGAGATCACCAACACACC
Val1127-Asn195-Arg426-Gly431	(1141) GACGGGGCAAGGAGATCACCAACACACC
Val1120-Thr202-Ile424-Ala433	(1087) GACGGGGCAAGGAGATCACCAACACACC
Leu122-Ser199-Arg426-Lys432	(1111) GACGGGGCAAGGAGATCACCAACACACC
Leu122-Ser199-Arg426-Gly431	(1111) GACGGGGCAAGGAGATCACCAACACACC
Lys121-Val200-Asn425-Lys432	(1099) GACGGGGCAAGGAGATCACCAACACACC
Val1120-Ile201-Ile424-Ala433	(1087) GACGGGGCAAGGAGATCACCAACACACC
Val1120-Ile201B-Ile424-Ala433	(1141) GACGGGGCAAGGAGATCACCAACACACC
Consensus	1171 1200
Leu122-Ser199 Tryp427-Gly431	(1141) GAGATCTTCGCCCCGGGGGGGGGACATG
Val1127-Asn195-Arg426-Gly431	(1171) GAGATCTTCGCCCCGGGGGGGGACATG
Val1120-Thr202-Ile424-Ala433	(1117) GAGATCTTCGCCCCGGGGGGGGACATG
Leu122-Ser199-Arg426-Lys432	(1141) GAGATCTTCGCCCCGGGGGGGGACATG
Leu122-Ser199-Arg426-Gly431	(1141) GAGATCTTCGCCCCGGGGGGGGACATG
Lys121-Val200-Asn425-Lys432	(1129) GAGATCTTCGCCCCGGGGGGGGACATG
Val1120-Ile201-Ile424-Ala433	(1117) GAGATCTTCGCCCCGGGGGGGGACATG
Val1120-Ile201B-Ile424-Ala433	(1117) GAGATCTTCGCCCCGGGGGGGGACATG

Consensus		(1171)	GAGATCTTCCGCCCGGGCGGGCATG
Leu122-Ser199	Tryp427-Gly431	(1171)	CGGCGCGAACTGGCZAGCGGAGCGCTAACAG
Val127-Asn195	Arg426-Gly431	(1201)	CGGGCGCAAGTGGCGCGCGCGCTAACAG
Val120-Thr202	Ile424-Ala433	(1147)	CGGGCGCGCGCGCGCGCTAACAG
Leu122-Ser199	Arg426-Lys432	(1171)	CGGGCGCAAGTGGCGCGCGCGCTAACAG
Leu122-Ser199	Arg426-Gly431	(1171)	CGGGCGCAAGTGGCGCGCGCGCTAACAG
Lys121-Val200	Asn425-Lys432	(1159)	CGGGCGCAAGTGGCGCGCGCGCTAACAG
Val120-Ile201	Ile424-Ala433	(1147)	CGGGCGCAAGTGGCGCGCGCGCTAACAG
Val120-Ile201B	Ile424-Ala433	(1147)	CGGGCGCAAGTGGCGCGCGCGCTAACAG
Consensus		(1201)	CGCGCGCAACTGGCGCAGCGAGCTGTACAG
Leu122-Ser199	Tryp427-Gly431	(1201)	TACAGCGGCGGCGGAGATCGAGGCCCTGGCG
Val127-Asn195	Arg426-Gly431	(1231)	TACAGCGGCGGCGGAGATCGAGGCCCTGGCG
Val120-Thr202	Ile424-Ala433	(1177)	TACAGCGGCGGCGGAGATCGAGGCCCTGGCG
Leu122-Ser199	Arg426-Lys432	(1201)	TACAGCGGCGGCGGAGATCGAGGCCCTGGCG
Leu122-Ser199	Arg426-Gly431	(1201)	TACAGCGGCGGCGGAGATCGAGGCCCTGGCG
Lys121-Val200	Asn425-Lys432	(1189)	TACAGCGGCGGCGGAGATCGAGGCCCTGGCG
Val120-Ile201	Ile424-Ala433	(1177)	TACAGCGGCGGCGGAGATCGAGGCCCTGGCG
Val120-Ile201B	Ile424-Ala433	(1177)	TACAGCGGCGGCGGAGATCGAGGCCCTGGCG
Consensus		(1231)	TACAAGGTGGTGAGGATCGAGGCCCTGGCG
Leu122-Ser199	Tryp427-Gly431	(1231)	GTGGCCCCCAGCGAGGCGAGCGCGCGCTG
Val127-Asn195	Arg426-Gly431	(1261)	GTGGCCCCCAGCGAGGCGAGCGCGCGCTG
Val120-Thr202	Ile424-Ala433	(1207)	GTGGCCCCCAGCGAGGCGAGCGCGCGCTG
Leu122-Ser199	Arg426-Lys432	(1231)	GTGGCCCCCAGCGAGGCGAGCGCGCGCTG
Leu122-Ser199	Arg426-Gly431	(1231)	GTGGCCCCCAGCGAGGCGAGCGCGCGCTG
Lys121-Val200	Asn425-Lys432	(1219)	GTGGCCCCCAGCGAGGCGAGCGCGCGCTG
Val120-Ile201	Ile424-Ala433	(1207)	GTGGCCCCCAGCGAGGCGAGCGCGCGCTG
Val120-Ile201B	Ile424-Ala433	(1207)	GTGGCCCCCAGCGAGGCGAGCGCGCGCTG
Consensus		(1261)	GTGGCCCCCAGCGAGGCGAGCGCGCGCTG
Leu122-Ser199	Tryp427-Gly431	(1261)	GTGCAGCGCGCGAGAGCGCGCGCGCTG
Val127-Asn195	Arg426-Gly431	(1291)	GTGCAGCGCGCGAGAGCGCGCGCGCTG
Val120-Thr202	Ile424-Ala433	(1237)	GTGCAGCGCGCGAGAGCGCGCGCGCTG
Leu122-Ser199	Arg426-Lys432	(1261)	GTGCAGCGCGCGAGAGCGCGCGCGCTG
Leu122-Ser199	Arg426-Gly431	(1261)	GTGCAGCGCGCGAGAGCGCGCGCGCTG
Lys121-Val200	Asn425-Lys432	(1249)	GTGCAGCGCGCGAGAGCGCGCGCGCTG
Val120-Ile201	Ile424-Ala433	(1237)	GTGCAGCGCGCGAGAGCGCGCGCGCTG
Val120-Ile201B	Ile424-Ala433	(1237)	GTGCAGCGCGCGAGAGCGCGCGCGCTG
Consensus		(1291)	GTGCAGCGCGCGAGAGCGCGCGCGCTG
Leu122-Ser199	Tryp427-Gly431	(1291)	GGCGCGCGCGCGCGCGCGCGCGCTG
Val127-Asn195	Arg426-Gly431	(1321)	GGCGCGCGCGCGCGCGCGCGCGCTG
Val120-Thr202	Ile424-Ala433	(1267)	GGCGCGCGCGCGCGCGCGCGCTG
Leu122-Ser199	Arg426-Lys432	(1291)	GGCGCGCGCGCGCGCGCGCGCTG
Leu122-Ser199	Arg426-Gly431	(1291)	GGCGCGCGCGCGCGCGCGCGCTG
Lys121-Val200	Asn425-Lys432	(1279)	GGCGCGCGCGCGCGCGCGCGCTG
Val120-Ile201	Ile424-Ala433	(1267)	GGCGCGCGCGCGCGCGCGCGCTG
Val120-Ile201B	Ile424-Ala433	(1267)	GGCGCGCGCGCGCGCGCGCGCTG
Consensus		(1321)	GGCGCGCGCGCGCGCGCGCGCTG
Leu122-Ser199	Tryp427-Gly431	(1291)	GGCGCGCGCGCGCGCGCGCGCTG
Val127-Asn195	Arg426-Gly431	(1351)	GGCGCGCGCGCGCGCGCGCGCTG
Val120-Thr202	Ile424-Ala433	(1297)	GGCGCGCGCGCGCGCGCGCGCTG
Leu122-Ser199	Arg426-Lys432	(1321)	GGCGCGCGCGCGCGCGCGCGCTG
Leu122-Ser199	Arg426-Gly431	(1321)	GGCGCGCGCGCGCGCGCGCGCTG

FIG. 5H

Lys121-Val1200-Asn425-Lys432	(1309) GGCGGCAGCACCATGGGCGCCGAGCCTG
Val1120-Ile201-Ile424-Ala433	(1297) GGCGGCAGCACCATGGGCGCCGAGCCTG
Val1120-Ile201B-Ile424-Ala433	(1297) GGCGGCAGCACCATGGGCGCCGAGCCTG
Consensus	(1351) GGCGGCAGCACCATGGGCGCCGAGCCTG
	1381 1410
Leu122-Ser199 Tryp427-Gly431	(1351) ACCCGACCGTCAGGGCCGCCAGCTGCTG
Val1127-Asn195-Arg426-Gly431	(1381) ACCCGACCGTCAGGGCCGCCAGCTGCTG
Val1120-Thr202-Ile424-Ala433	(1327) ACCCGACCGTCAGGGCCGCCAGCTGCTG
Leu122-Ser199-Arg426-Lys432	(1351) ACCCGACCGTCAGGGCCGCCAGCTGCTG
Leu122-Ser199-Arg426-Gly431	(1351) ACCCGACCGTCAGGGCCGCCAGCTGCTG
Lys121-Val1200-Asn425-Lys432	(1339) ACCCGACCGTCAGGGCCGCCAGCTGCTG
Val1120-Ile201-Ile424-Ala433	(1327) ACCCGACCGTCAGGGCCGCCAGCTGCTG
Val1120-Ile201B-Ile424-Ala433	(1327) ACCCGACCGTCAGGGCCGCCAGCTGCTG
Consensus	(1381) ACCCGACCGTCAGGGCCGCCAGCTGCTG
	1411 1440
Leu122-Ser199 Tryp427-Gly431	(1381) AGCGGCATCGTCACAGCAGAACACCTG
Val1127-Asn195-Arg426-Gly431	(1411) AGCGGCATCGTCACAGCAGAACACCTG
Val1120-Thr202-Ile424-Ala433	(1357) AGCGGCATCGTCACAGCAGAACACCTG
Leu122-Ser199-Arg426-Lys432	(1381) AGCGGCATCGTCACAGCAGAACACCTG
Leu122-Ser199-Arg426-Gly431	(1381) AGCGGCATCGTCACAGCAGAACACCTG
Lys121-Val1200-Asn425-Lys432	(1369) AGCGGCATCGTCACAGCAGAACACCTG
Val1120-Ile201-Ile424-Ala433	(1357) AGCGGCATCGTCACAGCAGAACACCTG
Val1120-Ile201B-Ile424-Ala433	(1357) AGCGGCATCGTCACAGCAGAACACCTG
Consensus	(1411) AGCGGCATCGTCACAGCAGAACACCTG
	1441 1470
Leu122-Ser199 Tryp427-Gly431	(1411) CTGGGGCCCATCGAGGCCAGCAGCACCTG
Val1127-Asn195-Arg426-Gly431	(1441) CTGGGGCCCATCGAGGCCAGCAGCACCTG
Val1120-Thr202-Ile424-Ala433	(1387) CTGGGGCCCATCGAGGCCAGCAGCACCTG
Leu122-Ser199-Arg426-Lys432	(1411) CTGGGGCCCATCGAGGCCAGCAGCACCTG
Leu122-Ser199-Arg426-Gly431	(1411) CTGGGGCCCATCGAGGCCAGCAGCACCTG
Lys121-Val1200-Asn425-Lys432	(1399) CTGGGGCCCATCGAGGCCAGCAGCACCTG
Val1120-Ile201-Ile424-Ala433	(1387) CTGGGGCCCATCGAGGCCAGCAGCACCTG
Val1120-Ile201B-Ile424-Ala433	(1387) CTGGGGCCCATCGAGGCCAGCAGCACCTG
Consensus	(1441) CTGGGGCCCATCGAGGCCAGCAGCACCTG
	1471 1500
Leu122-Ser199 Tryp427-Gly431	(1441) CTGAGACTGACCGTGTGGGCATCAAGCAG
Val1127-Asn195-Arg426-Gly431	(1471) CTGAGACTGACCGTGTGGGCATCAAGCAG
Val1120-Thr202-Ile424-Ala433	(1417) CTGAGACTGACCGTGTGGGCATCAAGCAG
Leu122-Ser199-Arg426-Lys432	(1441) CTGAGACTGACCGTGTGGGCATCAAGCAG
Leu122-Ser199-Arg426-Gly431	(1441) CTGAGACTGACCGTGTGGGCATCAAGCAG
Lys121-Val1200-Asn425-Lys432	(1429) CTGAGACTGACCGTGTGGGCATCAAGCAG
Val1120-Ile201-Ile424-Ala433	(1417) CTGAGACTGACCGTGTGGGCATCAAGCAG
Val1120-Ile201B-Ile424-Ala433	(1417) CTGAGACTGACCGTGTGGGCATCAAGCAG
Consensus	(1471) CTGAGACTGACCGTGTGGGCATCAAGCAG
	1501 1530
Leu122-Ser199 Tryp427-Gly431	(1471) CTGAGGCCCGCGCTGCTGGCGCTGGCGC
Val1127-Asn195-Arg426-Gly431	(1501) CTGAGGCCCGCGCTGCTGGCGCTGGCGC
Val1120-Thr202-Ile424-Ala433	(1447) CTGAGGCCCGCGCTGCTGGCGCTGGCGC
Leu122-Ser199-Arg426-Lys432	(1471) CTGAGGCCCGCGCTGCTGGCGCTGGCGC
Leu122-Ser199-Arg426-Gly431	(1471) CTGAGGCCCGCGCTGCTGGCGCTGGCGC
Lys121-Val1200-Asn425-Lys432	(1459) CTGAGGCCCGCGCTGCTGGCGCTGGCGC
Val1120-Ile201-Ile424-Ala433	(1447) CTGAGGCCCGCGCTGCTGGCGCTGGCGC
Val1120-Ile201B-Ile424-Ala433	(1447) CTGAGGCCCGCGCTGCTGGCGCTGGCGC
Consensus	(1501) CTGAGGCCCGCGCTGCTGGCGCTGGCGC
	1531 1560
Leu122-Ser199 Tryp427-Gly431	(1501) TACCTGAAGGACCAAGCAGCTGCTGGGCATC
Val1127-Asn195-Arg426-Gly431	(1531) TACCTGAAGGACCAAGCAGCTGCTGGGCATC

Val120-Thr202-Ile424-Ala433	(1477)	TACCTGAGCCACCGACGCCCTGGGGCATC
Leu122-Ser199-Arg426-Lys431	(1501)	TACCTGAGACGACCGACGCCCTGGGGCATC
Leu122-Ser199-Arg426-Gly431	(1501)	TACCTGAGACGACGCCCTGGGGCATC
Lys121-Val1200-Asn425-Lys432	(1489)	TACCTGAACTCCACGCCCTGGGGCATC
Val120-Ile201-Ile424-Ala433	(1477)	TACCTGAACTCCACGCCCTGGGGCATC
Val120-Ile201B-Ile424-Ala433	(1477)	TACCTGAACTCCACGCCCTGGGGCATC
Consensus	(1531)	TACCTGAGGGACCCAGCGCTGCTGGGCATC
	1561	1590
Leu122-Ser199 Tryp427-Gly431	(1531)	TGGGGCTGCAAGGGAAAGCTGATCTGCACCC
Val127-Asn195-Arg426-Gly431	(1561)	TGGGGCTGCAAGGGAAAGCTGATCTGCACCC
Val120-Thr202-Ile424-Ala433	(1507)	TGGGGCTGCAAGGGAAAGCTGATCTGCACCC
Leu122-Ser199-Arg426-Lys432	(1531)	TGGGGCTGCAAGGGAAAGCTGATCTGCACCC
Leu122-Ser199-Arg426-Gly431	(1531)	TGGGGCTGCAAGGGAAAGCTGATCTGCACCC
Lys121-Val1200-Asn425-Lys432	(1519)	TGGGGCTGCAAGGGAAAGCTGATCTGCACCC
Val120-Ile201-Ile424-Ala433	(1507)	TGGGGCTGCAAGGGAAAGCTGATCTGCACCC
Val120-Ile201B-Ile424-Ala433	(1507)	TGGGGCTGCAAGGGAAAGCTGATCTGCACCC
Consensus	(1561)	TGGGGCTGCAAGGGAAAGCTGATCTGCACCC
	1591	1620
Leu122-Ser199 Tryp427-Gly431	(1561)	ACCGCCGCTGCCCTGGAAAGGCAACCTGGACCC
Val127-Asn195-Arg426-Gly431	(1591)	ACCGCCGCTGCCCTGGAAAGGCAACCTGGACCC
Val120-Thr202-Ile424-Ala433	(1537)	ACCGCCGCTGCCCTGGAAAGGCAACCTGGACCC
Leu122-Ser199-Arg426-Lys432	(1561)	ACCGCCGCTGCCCTGGAAAGGCAACCTGGACCC
Leu122-Ser199-Arg426-Gly431	(1561)	ACCGCCGCTGCCCTGGAAAGGCAACCTGGACCC
Lys121-Val1200-Asn425-Lys432	(1549)	ACCGCCGCTGCCCTGGAAAGGCAACCTGGACCC
Val120-Ile201-Ile424-Ala433	(1537)	ACCGCCGCTGCCCTGGAAAGGCAACCTGGACCC
Val120-Ile201B-Ile424-Ala433	(1537)	ACCGCCGCTGCCCTGGAAAGGCAACCTGGACCC
Consensus	(1591)	ACCGCCGCTGCCCTGGAAAGGCAACCTGGACCC
	1621	1650
Leu122-Ser199 Tryp427-Gly431	(1591)	ATCAAGAGCCTGGACCGATCTGGACACAC
Val127-Asn195-Arg426-Gly431	(1621)	ATCAAGAGCCTGGACCGATCTGGACACAC
Val120-Thr202-Ile424-Ala433	(1567)	ATCAAGAGCCTGGACCGATCTGGACACAC
Leu122-Ser199-Arg426-Lys432	(1591)	ATCAAGAGCCTGGACCGATCTGGACACAC
Leu122-Ser199-Arg426-Gly431	(1591)	ATCAAGAGCCTGGACCGATCTGGACACAC
Lys121-Val1200-Asn425-Lys432	(1579)	ATCAAGAGCCTGGACCGATCTGGACACAC
Val120-Ile201-Ile424-Ala433	(1567)	ATCAAGAGCCTGGACCGATCTGGACACAC
Val120-Ile201B-Ile424-Ala433	(1567)	ATCAAGAGCCTGGACCGATCTGGACACAC
Consensus	(1621)	ATCAAGAGCCTGGACCGATCTGGACACAC
	1651	1680
Leu122-Ser199 Tryp427-Gly431	(1621)	ATGACCTGATGGCTGGGGCGCGNATC
Val127-Asn195-Arg426-Gly431	(1651)	ATGACCTGATGGCTGGGGCGCGNATC
Val120-Thr202-Ile424-Ala433	(1597)	ATGACCTGATGGCTGGGGCGCGNATC
Leu122-Ser199-Arg426-Lys432	(1621)	ATGACCTGATGGCTGGGGCGCGNATC
Leu122-Ser199-Arg426-Gly431	(1621)	ATGACCTGATGGCTGGGGCGCGNATC
Lys121-Val1200-Asn425-Lys432	(1609)	ATGACCTGATGGCTGGGGCGCGNATC
Val120-Ile201-Ile424-Ala433	(1597)	ATGACCTGATGGCTGGGGCGCGNATC
Val120-Ile201B-Ile424-Ala433	(1597)	ATGACCTGATGGCTGGGGCGCGNATC
Consensus	(1651)	ATGACCTGATGGCTGGGGCGCGNATC
	1681	1710
Leu122-Ser199 Tryp427-Gly431	(1651)	GCAACTACACCGCCCTGATCTCACCGTG
Val127-Asn195-Arg426-Gly431	(1681)	GCAACTACACCGCCCTGATCTCACCGTG
Val120-Thr202-Ile424-Ala433	(1627)	GCAACTACACCGCCCTGATCTCACCGTG
Leu122-Ser199-Arg426-Lys432	(1651)	GCAACTACACCGCCCTGATCTCACCGTG
Leu122-Ser199-Arg426-Gly431	(1651)	GCAACTACACCGCCCTGATCTCACCGTG
Lys121-Val1200-Asn425-Lys432	(1639)	GCAACTACACCGCCCTGATCTCACCGTG
Val120-Ile201-Ile424-Ala433	(1627)	GCAACTACACCGCCCTGATCTCACCGTG
Val120-Ile201B-Ile424-Ala433	(1627)	GCAACTACACCGCCCTGATCTCACCGTG
Consensus	(1681)	GCAACTACACCGCCCTGATCTCACCGTG

FIG. 5J

Leu122-Ser199 Tryp427-Gly431	(1711)	1711	1740
Val127-Asn195-Arg426-Gly431	(1711)	ATCGAAGGAGGCCAGAACCGACGGAG	
Val120-Thr202-Ile424-Ala433	(1657)	ATCGAAGGAGGCCAGAACCGACGGAG	
Leu122-Ser199-Arg426-Lys432	(1681)	ATCGAAGGAGGCCAGAACCGACGGAG	
Leu122-Ser199-Arg426-Gly431	(1681)	ATCGAAGGAGGCCAGAACCGACGGAG	
Lys121-Val1200-Asn425-Lys432	(1669)	ATCGAAGGAGGCCAGAACCGACGGAG	
Val120-Ile201-Ile424-Ala433	(1657)	ATCGAAGGAGGCCAGAACCGACGGAG	
Val120-Ile2018-Ile424-Ala433	(1657)	ATCGAAGGAGGCCAGAACCGACGGAG	
Consensus	(1711)	ATCGAAGGAGGCCAGAACCGACGGAG	
	1741		1770
Leu122-Ser199 Tryp427-Gly431	(1711)	ACGAGCAGGAGCTGCTGGAGCTGGACAG	
Val127-Asn195-Arg426-Gly431	(1741)	ACGAGCAGGAGCTGCTGGAGCTGGACAG	
Val120-Thr202-Ile424-Ala433	(1687)	ACGAGCAGGAGCTGCTGGAGCTGGACAG	
Leu122-Ser199-Arg426-Lys432	(1711)	ACGAGCAGGAGCTGCTGGAGCTGGACAG	
Leu122-Ser199-Arg426-Gly431	(1711)	ACGAGCAGGAGCTGCTGGAGCTGGACAG	
Lys121-Val1200-Asn425-Lys432	(1699)	ACGAGCAGGAGCTGCTGGAGCTGGACAG	
Val120-Ile201-Ile424-Ala433	(1687)	ACGAGCAGGAGCTGCTGGAGCTGGACAG	
Val120-Ile2018-Ile424-Ala433	(1687)	ACGAGCAGGAGCTGCTGGAGCTGGACAG	
Consensus	(1741)	ACGAGCAGGAGCTGCTGGAGCTGGACAG	
	1771		1800
Leu122-Ser199 Tryp427-Gly431	(1741)	TGGGCGAGCCTGTGGAACTGGTTGCRATC	
Val127-Asn195-Arg426-Gly431	(1771)	TGGGCGAGCCTGTGGAACTGGTTGCRATC	
Val120-Thr202-Ile424-Ala433	(1717)	TGGGCGAGCCTGTGGAACTGGTTGCRATC	
Leu122-Ser199-Arg426-Lys432	(1741)	TGGGCGAGCCTGTGGAACTGGTTGCRATC	
Leu122-Ser199-Arg426-Gly431	(1741)	TGGGCGAGCCTGTGGAACTGGTTGCRATC	
Lys121-Val1200-Asn425-Lys432	(1729)	TGGGCGAGCCTGTGGAACTGGTTGCRATC	
Val120-Ile201-Ile424-Ala433	(1717)	TGGGCGAGCCTGTGGAACTGGTTGCRATC	
Val120-Ile2018-Ile424-Ala433	(1717)	TGGGCGAGCCTGTGGAACTGGTTGCRATC	
Consensus	(1771)	TGGGCGAGCCTGTGGAACTGGTTGCRATC	
	1801		1830
Leu122-Ser199 Tryp427-Gly431	(1771)	ACGAGTGGCTGTGTTACATCAAGATCTC	
Val127-Asn195-Arg426-Gly431	(1801)	ACGAGTGGCTGTGTTACATCAAGATCTC	
Val120-Thr202-Ile424-Ala433	(1747)	ACGAGTGGCTGTGTTACATCAAGATCTC	
Leu122-Ser199-Arg426-Lys432	(1771)	ACGAGTGGCTGTGTTACATCAAGATCTC	
Leu122-Ser199-Arg426-Gly431	(1771)	ACGAGTGGCTGTGTTACATCAAGATCTC	
Lys121-Val1200-Asn425-Lys432	(1759)	ACGAGTGGCTGTGCTACATCAAGATCTC	
Val120-Ile201-Ile424-Ala433	(1747)	ACGAGTGGCTGTGTTACATCAAGATCTC	
Val120-Ile2018-Ile424-Ala433	(1747)	ACGAGTGGCTGTGTTACATCAAGATCTC	
Consensus	(1801)	ACGAGTGGCTGTGTTACATCAAGATCTC	
	1831		1860
Leu122-Ser199 Tryp427-Gly431	(1801)	ATCATGATCGTGGCGGCCTGGTGGCGCTG	
Val127-Asn195-Arg426-Gly431	(1831)	ATCATGATCGTGGCGGCCTGGTGGCGCTG	
Val120-Thr202-Ile424-Ala433	(1777)	ATCATGATCGTGGCGGCCTGGTGGCGCTG	
Leu122-Ser199-Arg426-Lys432	(1801)	ATCATGATCGTGGCGGCCTGGTGGCGCTG	
Leu122-Ser199-Arg426-Gly431	(1801)	ATCATGATCGTGGCGGCCTGGTGGCGCTG	
Lys121-Val1200-Asn425-Lys432	(1789)	ATCATGATCGTGGCGGCCTGGTGGCGCTG	
Val120-Ile201-Ile424-Ala433	(1777)	ATCATGATCGTGGCGGCCTGGTGGCGCTG	
Val120-Ile2018-Ile424-Ala433	(1777)	ATCATGATCGTGGCGGCCTGGTGGCGCTG	
Consensus	(1831)	ATCATGATCGTGGCGGCCTGGTGGCGCTG	
	1861		1890
Leu122-Ser199 Tryp427-Gly431	(1831)	CGCATCGTGTGACCGTGCTGACCACTGTG	
Val127-Asn195-Arg426-Gly431	(1861)	CGCATCGTGTGACCGTGCTGACCACTGTG	
Val120-Thr202-Ile424-Ala433	(1807)	CGCATCGTGTGACCGTGCTGACCACTGTG	
Leu122-Ser199-Arg426-Lys432	(1831)	CGCATCGTGTGACCGTGCTGACCACTGTG	
Leu122-Ser199-Arg426-Gly431	(1831)	CGCATCGTGTGACCGTGCTGACCACTGTG	
Lys121-Val1200-Asn425-Lys432	(1819)	CGCATCGTGTGACCGTGCTGACCACTGTG	

FIG. 5K

Val120-Ile201-Ile424-Ala433	(1807) CGCATCGTGTACCGTGTGAGCATCGTG
Val120-Ile201B-Ile424-Ala433	(1807) CGCATCGTGTACCGTGTGAGCATCGTG
Consensus	(1861) CGCATCGTGTACCGTGTGAGCATCGTG 1891 1920
Leu122-Ser199 Tryp427-Gly431	(1861) AACCGCGTGCAGCAGGCTACAGCCCCCTG
Val127-Asn195-Arg426-Gly431	(1891) AACCGCGTGCAGCAGGCTACAGCCCCCTG
Val120-Thr202-Ile424-Ala433	(1837) AACCGCGTGCAGCAGGCTACAGCCCCCTG
Leu122-Ser199-Arg426-Lys432	(1861) AACCGCGTGCAGCAGGCTACAGCCCCCTG
Leu122-Ser199-Arg426-Gly431	(1861) AACCGCGTGCAGCAGGCTACAGCCCCCTG
Lys121-Val1200-Asn425-Lys432	(1849) AACCGCGTGCAGCAGGCTACAGCCCCCTG
Val120-Ile201-Ile424-Ala433	(1837) AACCGCGTGCAGCAGGCTACAGCCCCCTG
Val120-Ile201B-Ile424-Ala433	(1837) AACCGCGTGCAGCAGGCTACAGCCCCCTG
Consensus	(1891) AACCGCGTGCAGCAGGCTACAGCCCCCTG 1921 1950
Leu122-Ser199 Tryp427-Gly431	(1891) AGCTTCAGACCCGCTCCCCCGCCCCCGCC
Val127-Asn195-Arg426-Gly431	(1921) AGCTTCAGACCCGCTCCCCCGCCCCCGC
Val120-Thr202-Ile424-Ala433	(1867) AGCTTCAGACCCGCTCCCCCGCCCCCGC
Leu122-Ser199-Arg426-Lys432	(1891) AGCTTCAGACCCGCTCCCCCGCCCCCGC
Leu122-Ser199-Arg426-Gly431	(1891) AGCTTCAGACCCGCTCCCCCGCCCCCGC
Lys121-Val1200-Asn425-Lys432	(1879) AGCTTCAGACCCGCTCCCCCGCCCCCGC
Val120-Ile201-Ile424-Ala433	(1867) AGCTTCAGACCCGCTCCCCCGCCCCCGC
Val120-Ile201B-Ile424-Ala433	(1867) AGCTTCAGACCCGCTCCCCCGCCCCCGC
Consensus	(1921) AGCTTCAGACCCGCTCCCCCGCCCCCGC 1951 1980
Leu122-Ser199 Tryp427-Gly431	(1921) GCCCCCGCCGCCCCGGCGCATCGAGGAG
Val127-Asn195-Arg426-Gly431	(1951) GCCCCCGCCGCCCCGGCGCATCGAGGAG
Val120-Thr202-Ile424-Ala433	(1897) GCCCCCGCCGCCCCGGCGCATCGAGGAG
Leu122-Ser199-Arg426-Lys432	(1921) GCCCCCGCCGCCCCGGCGCATCGAGGAG
Leu122-Ser199-Arg426-Gly431	(1921) GCCCCCGCCGCCCCGGCGCATCGAGGAG
Lys121-Val1200-Asn425-Lys432	(1909) GCCCCCGCCGCCCCGGCGCATCGAGGAG
Val120-Ile201-Ile424-Ala433	(1897) GCCCCCGCCGCCCCGGCGCATCGAGGAG
Val120-Ile201B-Ile424-Ala433	(1897) GCCCCCGCCGCCCCGGCGCATCGAGGAG
Consensus	(1951) GCCCCCGCCGCCCCGGCGCATCGAGGAG 1981 2010
Leu122-Ser199 Tryp427-Gly431	(1951) GAGGGCGCCGCCCCGGCGCATCGAGGAG
Val127-Asn195-Arg426-Gly431	(1981) GAGGGCGCCGCCCCGGCGCATCGAGGAG
Val120-Thr202-Ile424-Ala433	(1927) GAGGGCGCCGCCCCGGCGCATCGAGGAG
Leu122-Ser199-Arg426-Lys432	(1951) GAGGGCGCCGCCCCGGCGCATCGAGGAG
Leu122-Ser199-Arg426-Gly431	(1951) GAGGGCGCCGCCCCGGCGCATCGAGGAG
Lys121-Val1200-Asn425-Lys432	(1939) GAGGGCGCCGCCCCGGCGCATCGAGGAG
Val120-Ile201-Ile424-Ala433	(1927) GAGGGCGCCGCCCCGGCGCATCGAGGAG
Val120-Ile201B-Ile424-Ala433	(1927) GAGGGCGCCGCCCCGGCGCATCGAGGAG
Consensus	(1981) GAGGGCGCCGCCCCGGCGCATCGAGGAG 2011 2040
Leu122-Ser199 Tryp427-Gly431	(1981) AGCCCCCTGGTGCACGGCTCTGGCCCTG
Val127-Asn195-Arg426-Gly431	(2011) AGCCCCCTGGTGCACGGCTCTGGCCCTG
Val120-Thr202-Ile424-Ala433	(1957) AGCCCCCTGGTGCACGGCTCTGGCCCTG
Leu122-Ser199-Arg426-Lys432	(1981) AGCCCCCTGGTGCACGGCTCTGGCCCTG
Leu122-Ser199-Arg426-Gly431	(1981) AGCCCCCTGGTGCACGGCTCTGGCCCTG
Lys121-Val1200-Asn425-Lys432	(1969) AGCCCCCTGGTGCACGGCTCTGGCCCTG
Val120-Ile201-Ile424-Ala433	(1957) AGCCCCCTGGTGCACGGCTCTGGCCCTG
Val120-Ile201B-Ile424-Ala433	(1957) AGCCCCCTGGTGCACGGCTCTGGCCCTG
Consensus	(2011) AGCCCCCTGGTGCACGGCTCTGGCCCTG 2041 2070
Leu122-Ser199 Tryp427-Gly431	(2011) ATCTGGCAGCGCTCGCGAGCTGCTGCTG
Val127-Asn195-Arg426-Gly431	(2041) ATCTGGCAGCGCTCGCGAGCTGCTGCTG
Val120-Thr202-Ile424-Ala433	(1987) ATCTGGCAGCGCTCGCGAGCTGCTGCTG

Leu122-Ser199-Arg426-Lys432	(2011)	ATCTGGGACGACCTGCGCAGCCCTGCGCTG
Leu122-Ser199-Arg426-Gly431	(2011)	ATCTGGGACGACCTGCGCAGCCCTGCGCTG
Lys121-Val200-Asn425-Lys432	(1999)	ATCTGGGACGACCTGCGCAGCCCTGCGCTG
Val1120-Ile201-Ile424-Ala433	(1987)	ATCTGGGACGACCTGCGCAGCCCTGCGCTG
Val1120-Ile201B-Ile424-Ala433	(1987)	ATCTGGGACGACCTGCGCAGCCCTGCGCTG
Consensus	(2041)	ATCTGGGACGACCTGCGCAGCCCTGCGCTG
	2071	2100
Leu122-Ser199 Tryp427-Gly431	(2041)	TTCAGCTACCAACGGCTGCGCGACCTGATC
Val122-Asn195-Arg426-Gly431	(2071)	TTCAGCTACCAACGGCTGCGCGACCTGATC
Val120-Thr202-Ile424-Ala433	(2017)	TTCAGCTACCAACGGCTGCGCGACCTGATC
Leu122-Ser199-Arg426-Lys432	(2041)	TTCAGCTACCAACGGCTGCGCGACCTGATC
Leu122-Ser199-Arg426-Gly431	(2041)	TTCAGCTACCAACGGCTGCGCGACCTGATC
Lys121-Val200-Asn425-Lys432	(2029)	TTCAGCTACCAACGGCTGCGCGACCTGATC
Val1120-Ile201-Ile424-Ala433	(2017)	TTCAGCTACCAACGGCTGCGCGACCTGATC
Val1120-Ile201B-Ile424-Ala433	(2017)	TTCAGCTACCAACGGCTGCGCGACCTGATC
Consensus	(2071)	ATCTGGGACGACCTGCGCAGCCCTGCGCTG
	2101	2130
Leu122-Ser199 Tryp427-Gly431	(2071)	CTGATCGCCGCCCCATCGTGGAGCTGCTG
Val122-Asn195-Arg426-Gly431	(2101)	CTGATCGCCGCCCCATCGTGGAGCTGCTG
Val120-Thr202-Ile424-Ala433	(2047)	CTGATCGCCGCCCCATCGTGGAGCTGCTG
Leu122-Ser199-Arg426-Lys432	(2071)	CTGATCGCCGCCCCATCGTGGAGCTGCTG
Leu122-Ser199-Arg426-Gly431	(2071)	CTGATCGCCGCCCCATCGTGGAGCTGCTG
Lys121-Val200-Asn425-Lys432	(2059)	CTGATCGCCGCCCCATCGTGGAGCTGCTG
Val1120-Ile201-Ile424-Ala433	(2047)	CTGATCGCCGCCCCATCGTGGAGCTGCTG
Val1120-Ile201B-Ile424-Ala433	(2047)	CTGATCGCCGCCCCATCGTGGAGCTGCTG
Consensus	(2101)	CTGATCGCCGCCCCATCGTGGAGCTGCTG
	2131	2160
Leu122-Ser199 Tryp427-Gly431	(2101)	GGCCGCCGGGCTGGGGGGCCCTGAGATAC
Val122-Asn195-Arg426-Gly431	(2131)	GGCCGCCGGGCTGGGGGGCCCTGAGATAC
Val120-Thr202-Ile424-Ala433	(2077)	GGCCGCCGGGCTGGGGGGCCCTGAGATAC
Leu122-Ser199-Arg426-Lys432	(2101)	GGCCGCCGGGCTGGGGGGCCCTGAGATAC
Leu122-Ser199-Arg426-Gly431	(2101)	GGCCGCCGGGCTGGGGGGCCCTGAGATAC
Lys121-Val200-Asn425-Lys432	(2089)	GGCCGCCGGGCTGGGGGGCCCTGAGATAC
Val1120-Ile201-Ile424-Ala433	(2077)	GGCCGCCGGGCTGGGGGGCCCTGAGATAC
Val1120-Ile201B-Ile424-Ala433	(2077)	GGCCGCCGGGCTGGGGGGCCCTGAGATAC
Consensus	(2131)	GGCCGCCGGGCTGGGGGGCCCTGAGATAC
	2161	2190
Leu122-Ser199 Tryp427-Gly431	(2131)	TGGGGCAACCTGCTGCAAGTACTGGATCCAG
Val122-Asn195-Arg426-Gly431	(2161)	TGGGGCAACCTGCTGCAAGTACTGGATCCAG
Val120-Thr202-Ile424-Ala433	(2107)	TGGGGCAACCTGCTGCAAGTACTGGATCCAG
Leu122-Ser199-Arg426-Lys432	(2131)	TGGGGCAACCTGCTGCAAGTACTGGATCCAG
Leu122-Ser199-Arg426-Gly431	(2131)	TGGGGCAACCTGCTGCAAGTACTGGATCCAG
Lys121-Val200-Asn425-Lys432	(2119)	TGGGGCAACCTGCTGCAAGTACTGGATCCAG
Val1120-Ile201-Ile424-Ala433	(2107)	TGGGGCAACCTGCTGCAAGTACTGGATCCAG
Val1120-Ile201B-Ile424-Ala433	(2107)	TGGGGCAACCTGCTGCAAGTACTGGATCCAG
Consensus	(2161)	TGGGGCAACCTGCTGCAAGTACTGGATCCAG
	2191	2220
Leu122-Ser199 Tryp427-Gly431	(2161)	GAGCTGAAGAACAGGCCCTGAGCTGTC
Val122-Asn195-Arg426-Gly431	(2191)	GAGCTGAAGAACAGGCCCTGAGCTGTC
Val120-Thr202-Ile424-Ala433	(2137)	GAGCTGAAGAACAGGCCCTGAGCTGTC
Leu122-Ser199-Arg426-Lys432	(2161)	GAGCTGAAGAACAGGCCCTGAGCTGTC
Leu122-Ser199-Arg426-Gly431	(2161)	GAGCTGAAGAACAGGCCCTGAGCTGTC
Lys121-Val200-Asn425-Lys432	(2149)	GAGCTGAAGAACAGGCCCTGAGCTGTC
Val1120-Ile201-Ile424-Ala433	(2137)	GAGCTGAAGAACAGGCCCTGAGCTGTC
Val1120-Ile201B-Ile424-Ala433	(2137)	GAGCTGAAGAACAGGCCCTGAGCTGTC
Consensus	(2191)	GAGCTGAAGAACAGGCCCTGAGCTGTC
	2221	2250

FIG. 5M

Leu122-Ser199 Tryp427-Gly431	(2191) GACGCCATCGCCATCAGCGTGGCCGAGGGC
Val1127-Asn195-Arg426-Gly431	(2221) GACGCCATCGCCATCAGCGTGGCCGAGGGC
Val1120-Thr202-Ile424-Ala433	(2167) GACGCCATCGCCATCAGCGTGGCCGAGGGC
Leu122-Ser199-Arg426-Lys432	(2191) GACGCCATCGCCATCAGCGTGGCCGAGGGC
Leu122-Ser199-Arg426-Gly431	(2191) GACGCCATCGCCATCAGCGTGGCCGAGGGC
Lys121-Val200-Asn425-Lys432	(2179) GACGCCATCGCCATCAGCGTGGCCGAGGGC
Val1120-Ile201-Lle424-Ala433	(2167) GACGCCATCGCCATCAGCGTGGCCGAGGGC
Val1120-Ile201B-Ile424-Ala433	(2167) GACGCCATCGCCATCAGCGTGGCCGAGGGC
Consensus	(2221) GACGCCATCGCCATCAGCGTGGCCGAGGGC
	2251 2280
Leu122-Ser199 Tryp427-Gly431	(2221) ACCGACCCGATCATCGAGGTGGCCCAAGCGC
Val1127-Asn195-Arg426-Gly431	(2251) ACCGACCCGATCATCGAGGTGGCCCAAGCGC
Val1120-Thr202-Ile424-Ala433	(2197) ACCGACCCGATCATCGAGGTGGCCCAAGCGC
Leu122-Ser199-Arg426-Lys432	(2221) ACCGACCCGATCATCGAGGTGGCCCAAGCGC
Leu122-Ser199-Arg426-Gly431	(2221) ACCGACCCGATCATCGAGGTGGCCCAAGCGC
Lys121-Val200-Asn425-Lys432	(2209) ACCGACCCGATCATCGAGGTGGCCCAAGCGC
Val1120-Ile201-Ile424-Ala433	(2197) ACCGACCCGATCATCGAGGTGGCCCAAGCGC
Val1120-Ile201B-Ile424-Ala433	(2197) ACCGACCCGATCATCGAGGTGGCCCAAGCGC
Consensus	(2251) ACCGACCCGATCATCGAGGTGGCCCAAGCGC
	2281 2310
Leu122-Ser199 Tryp427-Gly431	(2251) ATCCGCGCCGCCCCCTCCATCAGCATCCCCCGC
Val1127-Asn195-Arg426-Gly431	(2281) ATCCGCGCCGCCCCCTCCATCAGCATCCCCCGC
Val1120-Thr202-Ile424-Ala433	(2227) ATCCGCGCCGCCCCCTCCATCAGCATCCCCCGC
Leu122-Ser199-Arg426-Lys432	(2251) ATCCGCGCCGCCCCCTCCATCAGCATCCCCCGC
Leu122-Ser199-Arg426-Gly431	(2251) ATCCGCGCCGCCCCCTCCATCAGCATCCCCCGC
Lys121-Val200-Asn425-Lys432	(2239) ATCCGCGCCGCCCCCTCCATCAGCATCCCCCGC
Val1120-Ile201-Ile424-Ala433	(2227) ATCCGCGCCGCCCCCTCCATCAGCATCCCCCGC
Val1120-Ile201B-Ile424-Ala433	(2227) ATCCGCGCCGCCCCCTCCATCAGCATCCCCCGC
Consensus	(2281) ATCCGCGCCGCCCCCTCCATCAGCATCCCCCGC
	2311 2340
Leu122-Ser199 Tryp427-Gly431	(2281) CGCATCGCCAGGGCTTCGAGCGGGCCCTG
Val1127-Asn195-Arg426-Gly431	(2311) CGCATCGCCAGGGCTTCGAGCGGGCCCTG
Val1120-Thr202-Ile424-Ala433	(2257) CGCATCGCCAGGGCTTCGAGCGGGCCCTG
Leu122-Ser199-Arg426-Lys432	(2281) CGCATCGCCAGGGCTTCGAGCGGGCCCTG
Leu122-Ser199-Arg426-Gly431	(2281) CGCATCGCCAGGGCTTCGAGCGGGCCCTG
Lys121-Val200-Asn425-Lys432	(2269) CGCATCGCCAGGGCTTCGAGCGGGCCCTG
Val1120-Ile201-Ile424-Ala433	(2257) CGCATCGCCAGGGCTTCGAGCGGGCCCTG
Val1120-Ile201B-Ile424-Ala433	(2257) CGCATCGCCAGGGCTTCGAGCGGGCCCTG
Consensus	(2311) CGCATCGCCAGGGCTTCGAGCGGGCCCTG
	2341 2352
Leu122-Ser199 Tryp427-Gly431	(2311) CTGTAACCTCGAG
Val1127-Asn195-Arg426-Gly431	(2341) CTGTAACCTCGAG
Val1120-Thr202-Ile424-Ala433	(2287) CTGTAACCTCGAG
Leu122-Ser199-Arg426-Lys432	(2311) CTGTAACCTCGAG
Leu122-Ser199-Arg426-Gly431	(2311) CTGTAACCTCGAG
Lys121-Val200-Asn425-Lys432	(2299) CTGTAACCTCGAG
Val1120-Ile201-Ile424-Ala433	(2287) CTGTAACCTCGAG
Val1120-Ile201B-Ile424-Ala433	(2287) CTGTAACCTCGAG
Consensus	(2341) CTGTAACCTCGAG

FIG. 5N

SEQ ID NO:3 VAL120-ALA204

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGTGAGCA
 GTCTCTGTTGCCCAGGCGTGGAGAAGCTGTGGTGACCGCTGACTACGGCTGCCGTG
 TGGAAAGGAGGCCAACACCAACCTGTCTGCGGCCAGGGACGCGCAAGGCCCTACGACACGGAGT
 GCACAAACCTGTGGGCCACCCACCGCTGCGTGGCAACGACCCCACCCCAAGGAGATCTGTGCT
 GGAGAACGTGACCGAGAACTTCACATGTGGAAGAACAAACATGTGGAGCAGATGACAGAG
 GACATCATCAGCCTGTGGGACCAAGGCGCTGCGTGGGCCGGCGCTGCCATCTGAAAGT
 GGTGAGCTTCAAGGCCATCCCCATCTCAACTCTGCGCCCCCGCCGCTTGCACATCTGAAAGT
 CAACGACAAGAAAGTTCAACGGCAAGGGCCCTGACCAACGTGAGCAGCCTGAGCTGACCC
 ACGGCATCGGCCCTGGTGGACACCCAGCTGCTGTAACGGCAAGCTGGCAAGGGGG
 GTGGTGTACCGCAGGAGAACTTCACCGACAACGCCAAGACCATCATGTGAGCTGAAGGA
 GAGCGTGGAGATCACTGACCCCCAACACACACCCGCAAGAGCATACCATCGGCC
 CCGGGCGGCCCTTCAACGCCACGGCGACATCTGGCGACATCGCCAGGCCACTGCAACA
 TCAAGCGGGAAGAATGGAACAAACCCGTGAAGCAGATCTGTAACAGGCTGCCAGITC
 GGCAACAAAGACCATGTTCAAGCAGACGGCGGGAGCCCGAGATGTTGATGACAG
 CTTCACACTGGCGCGCGAGTTCTACTGCAACAGCACCCAGCTGTTAACAGCACCTGAA
 CAAACACCATCGGCCCAACACCAAACGGCAACATACCCCTGCCCTGCGCATACAGAGA
 TCATCGACCGCTGGCAGGAGTTGGCAAGGCCATGTACGCCCCCCCATCGCGGAGATC
 CGCTGAGCAGAACATACCGGCCCTGCTGTAACGGCGACGGGGCAAGGAGATCAGCAA
 CACCAAGGAGATCTTCCGCCGGGGCGGCGACATGCGCAGAACACTGGCGCAGGAGCTG
 ACAAGTACAAGGTGGTAAGATGGACGCCCTGGCGTGGGCCACCAAGGCCAAGGCCG
 GTGGTGTACGGCGGAGAACGGCGCGTGAACCTTGGCGGCGCATGTTCTGGGCTTCCGGGCC
 GCGCGCAACCATGGGCCCGCAACCTGACCTCTGACGGCGTCAAGGCCCGAGCTGCTGAG
 CGCATCTGTCAGCAGCAGAACACCTGCTGCGGCCATCGAGGCCAGCAGCACTGCTG
 AGCTGACCGTGTGGGCATCAACGAGCTGCAAGGGCGGTGCTGCCGTGGAGCGTACCTG
 AAGGACACGACGCTGCTGGCATCTGGGCTGCAAGCGCAAGCTGATCTGCAACCCGGCT
 GCCCTGAAACGCCAGCTGGAGCAACAGAGCTGACCGAGATCTGAAACACATGACTTGA
 TGGAGTGGAGGGAGGAGATGACAACACTACCAACCTGATCTACACCCCTGATCTGAGGGAG
 CAGAACACGAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGAACAGTGGGCCAGCGT
 GGAACCTGTTGACATCAGCAAGTGGCTGTTGTAACATCAAGATCTTACATGATCTGIGGGCG
 GCGCTGGGCGCTGCGCATCTGTTGACCGCTGCTGAGCATCTGTAACCGCGTGGCGCAGGGCT
 ACAGCCCCCTGAGCTTCAAGACCCGCTTCCCCGCCCGGGCCGAGGGCA
 TCGAGGGAGGAGGGCGGGAGCAGCGAAGCGAGCAGGCCCTGTGCAAGGCCCTGCTG
 GCCCTGATCTGGGACGACCTGGCGACGCCCTGCTGCTGCTGCTGCTGCAAGGCCCTG
 ATCTCTGATGCCGCCGACATCTGTTGAGCTGCTGGGCCGGCGCTGGGAGGGCGCTGAAGTAC
 TGGGGCAACCTGCTGAGTACTGGATCAAGAGCTGAAAGAACAGGCCGTGAGCTTCTGCA
 CGGCATGCCATCGCGTGGCGAGGGCACCGACCGCATCTGAGGTGGCCAGCGCATG
 GCGCGCCCTTCTGACATCCCCGCCGATCGGCCAGGGCTTCTGAGCGCGCCCTGCTGAA
 TCGAG

FIG. 6

SEQ ID NO:4 VAL120-ILE201

FIG. 7

SEQ ID NO:5 VAL120-ILE201B

GAATTGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGCTGCTGTGTTGGAGCAGTCCTCGTTCCAGGCCGCGCGTGGAGAAAGCTGTGGGTGACCGTGTACTAGGGGTGCCCCGTGCTGGAAGGAGGCCA
 CCAACACCCCTGTTGCGAGCAGGCCAACGGCTTACAGACACCGAGGTGCAACAGCTGTGAGAACGATGACCGAGAACGAGATCTACAGCTGTTGGAGACGCCA
 AGGCCCTGGTGGCCCAACGGACCCCAACGGCAAGAGATGCTGCTGAGAACGAGATCTACAGCTGTTGGAGACGCCA
 TGTGGAAGAACAACTATGGTGGAGCAATGCAAGGAGACATCATCAGCCTGTGGAGACGCCAAGCTGAGCTGGAGACGCCA
 CCTCTGGTGGCCGACATACCCAGGCCGCCCCAAGGTGAGCTTCAGGCCATCCATCACTACTCGCC
 CCCCGCGCGCTTGCATACCCAGGCCGCCCCAAGGTGAGCTTCAGGCCATCCATCACTACTCGCC
 GAGCACCGTGTGACGTGCAACCCACGGCATTCGGCCCGTGTGGAGACCCAGCTGCTGTAACGCCAGCCT
 GGGCAAGGGGGCTGGTGTGATCGCAGGGAGAACCTACGGCAACAGGCAAGACCATATGCTGAGCT
 GAAGGAGAGGGTGGAGATCACTGCAACCCGCCAACAACAAACACCCGCAAGAGCATACCATGGCGCT
 CGGCCGCGCCTCTACGCCACGGCGACATCATCGCGACATCCGCCAGGCCACTGCAACATCAAGGGC
 GAGAACGGTGAACACACCCCTGAAGCAGATCGTGAACAGCTGCAAGGCCAGGGCCACAAAGACCATC
 GTGTTCAAGCAGACAGCACCCGGCTGAGATGTTGATGCAACAGCTTCAACTCGGGGGGGAGTTTC
 TCTACTGCAACAGCACCCGGCTGAGATGTTGATGCAACAGCACCCGGCTGAGAACACCCACCAAC
 GGCCACCATACCCCTGGCCCTGACCGTCAACGACACTGGAGAACACCATCGGCCCAACAAAGACCATG
 TACGGGGCCCGCCCGAGATCCGAGCACCAACTACCGGCCCTGGTGTGACCGCCAGCAG
 GCGGCAAGGGAGATCAGAACACCCAGGAGATCTTCCGGCCCGGCCGGCGCGACATGGCGCAACATGGC
 GCAAGCAGCTGTAACGGCAAGCTGAGATCTGCAACGGCCCGAGGGCCCTGGGCCACAAAGGGCAAGC
 GCGCCGCGTGTGCAAGGGCGAGAAGGGCGCCGCGCTGAGCTTCCAGGCCCTGAGCTGACCCCTGAGCTG
 CGGCAGCACCATGGGCCCGCCGCGCTGAGCTTCCAGGCCCTGAGCTGACCCCTGAGCTGAGCCGCGATCGT
 GCAGCAGCAGAACACCCCTGCGCCGCGCCATCGAGGCCCTGAGCTGAGCCCGCAGCTGAGCTGAGCCGCGATCGT
 CATCAAGCAGCTGCAAGGGCCCGTGTGCGCCGCGTGGAGGACGCTAACCTGAAAGGAGACAGCAGCTGAGCTGTTGG
 CTGGGGCTGCAACGGCGCAAGCTGAGTCTGCAACACCCGGCGTGGAGGCCCTGGAGAACCCAGCTGGAGACAAAGAG
 CTGGCAACAGATCTGAAACACATGAACTGGATCTGGATGGAAGTGGAGATGCAACACTACACCAACCT
 GATCTACACCCCTGATCGAGGGAGAGCCAGAACACAGCAGGAGAAAGAAGAACAGAAGGAGCTGCTGGAGCTGG
 ACAGTGGGCCAGCTGTGAACTGGTGTGACATCAGCAAGTGGTGTGTTGATCACAAAGATCTCATCAT
 GATCTGGGGCCGCGCTGGTGGCCCTGGCCATCGTGTGACCGTGTGAGACATCTGAAACCCGGCTGGGCCAG
 GGGTACAGCCCCCTGAGCTTCCAGACCCGTGACCGCCCGGCCGCGGCCACCGCCCCAGGGCATCG
 AGGAGGGAGGGCGCCGAGCGCGACCCGACCCGCAAGCAGCAACGCCCTGGTGTGACGGCCCTGCTGGCCCTGATCT
 GGGAGCAGCTGGCGCAAGCTGCGCTGAGCTTACAGTACCAACGGCTGCGGCCAGCTGAGCTGACGGCCCG
 CATCGTGGAGCTGCTGGCCGCCGCGCTGGAGGCCCTGGAGAACCTGCTGAGCTGAGTACTG
 GATCCAGGAGCTGAGAACACGGCGCTGAGCTGTTGACGGCCATGCCACATGCCCTGGCCAGGGCAC
 CGACCGCATCGAGGTTGGCCAGGCCATGCCCTGGCCATGCCACATCCCCGCCGCGCATCGGCC

FIG. 8

SEQ ID NO:6 LYS121-VAL200

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
GTCTCGTTGCCAGCGCCGTGAGAAGCTGTGGTACCGTGTACTACGGCTGCCGTG
TGGAAAGGAGGCCAACACCAACCCCTGTTCTGCCAGGGACGCCAACGGCTACGACACCGAGGT
GCACAAACGTGTGGGCCACCCACGCCCTGCGTGTGCCAACGCCAACCCCCAAGGAGATCGTGT
GGAGAACGTGACCGAGAACACTTCAACATGTGGAAGAACACATGGTGGAGCAGATGACAGAG
GACATCATCAGCTGTGGAGACCGAGGCCCTGCGTGTGAAGGCCCGCTGATCACCCA
GGCTGCCCGAACGGTGAAGCTTGTGAGGCCCATCCCAACTACTGCCCCCCCGCCGCTTGC
CATCTGAAAGTGAACGACAAGAAGTTCAACGGCAGGGCCCTGACCAAAACGTGACCCAG
TGCAGTGCACCCCGCATCCCCCTGTTGAGCAGGGACACTTACCGACACGCCAACCATCATCGT
CAGCTGAAGGAGAGGGCTGGAGATCACTGACCCGGCCCAACAAACACACCCGAAGAGCAT
CACCATGGGCCCGCCGCCGCTTCTAGGCCACCGGCCGACATCATCGCGACATCGGCCAGGC
CAACTGAAACATCAGCGGGAAAGTGAACACCCCTGAAAGCAGATCGTGAACAACTG
AGGGCCAGTTGGCAAAAGACCATGTTGAGCAGGGCAGGGCGGGAACCCGAAGATC
GTGATGACAGCTTCACTCGCGCGCGAGTTCTTACTGCAACAGCACCCAGCTGTTAAC
AGCACCTGGAAACACCATCGGCCCAACAAACACCGGCCACATCACCCCTGCCCTGGC
CATCAAGCAGATCATCAACCGCTGGCAGGAGTTGGCAAGGCCCATGTAGCCCCCCCCATCC
GGGCCAGATCGCTGCGACGAAACATCACCGGCCCTGCTGTCAGCCGCAGGGCGGAAG
GAGATCGCAACACCCAGGAGATCTCCGCCCGGGGGGAGCATCGCGACAACTGGCG
CAGCGAGCTGTAACAGTAAAGGTGGTGAAGATGAGGCCCTGGGGTGGCCCAACAAAG
CAACAGGCCCGCTGGTGTGAGCAGGGAGAAGGGCCGCTGACCCCTGGGGCCGATGTTCTGGC
TTCTGGGCCCGCCGGACACCATGGGGCCGGCAGCTGACCTGACCGTGAAGGCCCG
CAGCTGCTGAGCGGACATCGTCAGCAGCAGAACACCTGCTGCGGCCATCGAGGGCCAGCA
GCACCTGCTGAGCTGACCTGGCTGGGGCATCAAGCAGCTGCGAGGGCCCGCTGCTGGCGTGG
AGGGCTACCTGAAGGACACGACGACTGCTGGCATCTGGGCTGAGCGGCCAAGCTGATCTGC
ACCAACCCGCTGGCTTGAACCCAGCTGGAGCAAAAGAGCCTGGACCCAGATCTGGAAACAA
CATGACCTGGATGGAGTGGAGGGGGAGATGACAACTACACCAACCTGATCTACACCTGA
TCGAGGAGAGCCAGAACAGCAGGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTG
GGCCAGCTGGAAGTGGTTCGACATCAGCAAGTGGCTGTTGATCATCAAGATCTCATCAT
GATCTGTTGGCCGGCTGGCTGGGCTGCGCATCTGTTGACCGCTGCTGAGCATCTGTAACCGCGT
GGCCAGGGCTACAGCCCCCTGAGCTTCCAGAACCCGCTTCCCCGCCCGGGCCCGACCG
CCCCGAGGGCATCGAGGAGGGGGGGGGAGCGGCCAGCGCAGCCGAGGCCCTGTGTC
ACGGCTGCTGGCCCTGATCTGGAGCACCTGGCAGCCCTGTCAGCTGCTTGTGAC
TGGCGACCTGATCTGATCTGGAGGCCCATGTGGAGGAGCTGCTGGGGCCGGCTGGAGG
CCTGAAGTACTGGCAACCTGCTGAGTACTGGAGTCCAGGAGCTGAAAGAACAGGGCTG
AGCTGTTGAGCCATGCCATGCCGTGGCCAGGGCACCGACCGCATCTCGAGGGTGGCC
CAGCGCATCGCCCGGGCTTCTGACATCCCCCGCCATCGCCAGGGCTTCGAGCGCC
CTGCTGAACTCGAGCGTGT

FIG. 9

SEQ ID NO:7: LEU122-SER199

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCCTCTGTTTCGCCCAGGGCGTGGAGAAAGCTGTTGGGTGACCGTGTACTACGGCTGCCCCGTG
 TGGAAAGGAGGCCACCCACCCACCTGTCGCGCAGCAGCAGCAAGGCCCTACGACACCGAGGT
 GCACAACTGTTGGGCCACCCACCCCTGCGTGGCCAGGCCAACCCCAAGGAATGCTGCTGCT
 GGAGAACCTGAGCCGAGAACCTCAACATGGTGAAGAACAAACATGTGGAGCAGATCAGCAG
 GACATCATCAGCTGTGGGACAGGCTGAAGGCTGCGTGAAGCTGGGCAACAGCTGAT
 CACCCAGGCCCTGCCCAAGGTGGCTGACGCCATCCCCATCCACTACTGGGCCCGCCCG
 CTTCGCACTCTGAAGTGCACAGACAAGAATTCAGGCCACGGCGACGGGCCCTGCAACAACTGTA
 GCACCGTGCACTGGCACCCAGGCCATCCGGCCCTGGTGAACCCAGCTGCTGACAGCGC
 AGGCTGGAGGCCAGGGCGTGGTGAATCGCAGCAGGAAGAACCTACCCGACAAAGGCCAACAGCAT
 CATCTGTCAGCTGAAGGAGCCTGGAGATCAACTGCAACCCGCCAACAAACAAACCCGCA
 AGAGCATCACCATCGCCCCGGCGGCCTTACGCAACGGCGACATCATGGGACATCC
 GCGAGGAGAACATCGGGCAGAACATGGGGCAGAAGTGGAAAGAACAAACCCCTGAAGCAGATCTGACC
 AAGCTGAGGCCAGTTGGCAACAGACATCTGTTCAAGCAGAGCAGGGGGGGGACCC
 CGAGATCTGTAAGCTCACGCTTCACTGGGGGGAGATCTCTTACTGCAACAGCACCCAGCT
 GTTCAACAGCAGCTGGAAACAAACCATCGGCCAACAAACCAAGGCCATCATCTG
 CCTGGCCGATCAAGCAGATCATCACCGCTGGCAGGGAGTGGCAAGGCCATGTAAGCCCCC
 CCCATCCGGCCAGATCGCGCTGCAAGCAACATCACGGGCTCTGCTGACGCCGGCGAC
 GGCAAGGAGATCACGCAACACCCAGAGATCTTCCGGGGGGGGGGCGCGACATCGGGACAA
 CTGGCAGCAGCTGTAAGTCAAGGAGTGGTGAAGATCTGAGGCCCTGGCGTGGCCCCCA
 CCAAGGGCAAGGCCGGCTGGTGCAGGGCGAGAAGGGGGCCGTGACCCCTGGGCGCATTTG
 CTGGGCTTCTGGGGCCGGCGCAGCACCATTGGGGCCGCGAGCTGACCCCTGGGCGCAT
 GCGGGCAGCTGCTGAGGCCATCTGAGCAGCAACCTGGTGTGGCGCCATCGAGGC
 CCAGCAGCAGCTGCTGAGCAGCTGCTGGGAGTCAAGCAGCTGCAAGGCCGGCTGCTGG
 CGTGGAGCCTAATCTGAGGACAGCAGCTGCTGGGAGTCTGGGAGTCAAGCAGCTG
 ATCTGACCAACCCGGCTGAGCTGGGAGACGCCAGCTGGAGCAACAAAGAGCTGGACAGATCTG
 GAACAAACATGACCTGGATGGAGTGGAGCTGGAGCGAGAATGCAACACTAACCAACCTGATCTACA
 CCTGTGATCGAGGAGAGCCAGAACCCAGCAGGAGAAGAACGAGCAGGGAGCTGAGCTGG
 CAAGTGGGGCAGCTGTGGAACCTGTTGCAATCATCAGCAATGGCTGTGTTACATCAAGATCTT
 CATCATGATCTGGGGGGCTGGGGGCTGCGCATCTGTTACCTGCTGAGCATCTG
 CCGGGTGGCCAGGGCTACAGCCCTGAGCTTCCAGACCCGGCTTCCCGGCCCGGGCC
 CGACCGCCCCGAGGGCATCGAGGAGGGGGCGGAGCGCCAGCCGACCGCAAGGCC
 CTGGTGCACGGGCTGCTGGCCCTGATCTGGGAGCAGCTGGCAGCCTGTGCTGTCAGCTAC
 CACGGCCCTGGCGGAGCTGATCTGAGCTGGGGGGGGGGGGGGGGGGGGGGGGGG
 TGGGAGGCCCTGAAGTACTGGGGCAACCTGCTGAGCTACTGGATCCAGGAGCTGAAGAACAG
 CGCGCTGAGCCTGTTGCAAGCCATCGCCATCGCCAGGGGACCCGACCGCATCTCGA
 GGTGCCCCAGGCCATCGGGCGGCCGCTTCCGACATCCCCCGCCGACATCGGCCAGGGCTTCGA
 CGCGCCCTGCTGAACTCGAGCGTGT

FIG. 10

SEQ ID NO:8 VAL120-THR202

GAATTCGCACCATTGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTTCGTTTCTGCCAGCGCGTGGAGAACGCTGTTGGTGACCGTGTACTACGGCGTGCCCGTG
 TGGAGGAGGGCCACCAACCCCTTGCTCGCGCAAGGCCAACGACACCAGGT
 GCACAAACGTTGGGCCACCCACGGCTGCGTGGCCACCGAACCCAAACCCCAAGGAGATCGTGTCT
 GGAGAACGTCACCGAGAACCTTCAACATGTGAAAGAACAACTGTGGAGCAGATGCACGAG
 GACATCATCAGCTGTGGGACAGAGCCTGAAGGCCCTGCGTGGGGCGGCCACCCAGGCC
 CCCAAAGGTGAGCTTCGAGGCCATCCCACTACTCGGCCCCCGGCCCTTCGCCATCT
 GAAAGTGCACCGACAAGAAGTTCACAGCGACGGCCAGCTGCTGTAACGGCAGCTGGCCAG
 GCACCCACGGCATCCGGCCCTGTTGAGCACCAGCTGCTGCTGTAACGGCAGCTGGCCAG
 GAGGGCGTGGTGTGGCGAGGAACCTTACCGAACAGCGAACACCATCATGTGAGCT
 GAAAGGAGAGCTGGAGATCACTGCACCCGCCAACAAACACCCGAAGGAGATCACCA
 TCGGGCGGCCCTCTACCGGAGACATCGGGGACATCGGGGACATCGGGGAGCCACT
 GCACACATCAGGGCGAGAACAGTGGAAACACCCCTGAAGCAGATGTGACCAAGCTGAGGCC
 CAGTTGCGCAACAAAGACCATCTGTTCAAGCAGAGCAGCGGGGGACCCCGAGATCTGTGAT
 GCACAGCTTACAGCGGGGGAGTTCTACTGCAACAGCACCCAGCTGTTCAACAGC
 CTGGAACACACCATCGGCCAACAAACCAACCGCACCATCACCTGCCCTGCCGATCA
 AGCAGAGTCATCACCGCTGGAGAGTGCGGAGACGGCATGACGCCCTCCCATCGCGC
 CAGATCGGCTGCAGCAGAACATCACCGGCCCTGCTGTCGACCCGGACGGCGGAAAGGAGAT
 CAGCAACACCAACCGAGATCTTCCGCCCGGCCGGGCGACATGCCGACAATGGCGCAGCG
 AGCTGTACAAGATGGTGGTGAAGATCGAGGCCCTGGGGCTGCCCCCAGGAGGCG
 CGCCCGCTGGTGCAGCGGGAGAACGGCGGGCTGACCCCTGGGGCCATGTTCTGGCTTCC
 GCGCGCCCGCAGCACCATGGCGCCAGCAGCTGACCCCTGACCGTGAAGGGCCGCG
 GCTGTGAGGGCATCGTGCAGCAGAACACCTGACGGCACATCGAGGCCATCGAGGCC
 TCGAGCGAGCTGACCGTGTGGGCATCAAGCAGCTGCGAGGCCGCTGCTGGCGTGGAGGC
 TACCTGAAGGACACAGCAGCTGCTGGGCTCTGGGGCTGAGCGGGCAAGCTGATCTGCCAAC
 CGCCGTGGCTGGAGCGCAGCTGGAGAACAAAGAGCTGGACCATGTTGAAACACATGA
 CCTGGATGGAGTGGAGGGCGAGATGCAACACTACACCAACTGATCTACACCCCTGATCGAG
 GAGAGCCAGAACACAGCAGGGAGAACAGCAGGGAGCTGCTGGAGCTGAGAATGGCCA
 GCGCTGGGAACATGGTGTGACATCACAGCAAGCTGCTGTGATCATCAAGATCTCATGATG
 TGGCGCCCTGGTGGCTGGCATGTTACCGTGTGACATGGTGAACCGCGCTGGCC
 AGGGCTACAGCCCCCTGAGCTTCAGACCCGCTTCCGCCCGGGGCCCGACCCCC
 AGGGCATCGAGGAGGGGGGGAGCGCGACGGGAGCCGAGCAGAGCCCCCTGGTGCACGG
 CCTGCTGGCCCTGATCTGGGACGACCTGCGCAGCTGTGCTGTGCTGCTG
 GACCTGATCTGATCGCCGCCCATGTTGAGGCTGCTGGGGCCGGCGTGGAGGCC
 GAAGTACTGGGGCAACCTGCTGCAAGTACTGGAGTCAAGGAGCTGAAAGAACAGGCC
 TGTGAGGGCATGCGCATCGCCCTGGGCCAGGGCACCGCAGCTGACCC
 GCATCGGCCGCCCTCTGCACATCCCCGCCGATCGGCCAGGGCTCGAGCGGCC
 GTGAACGCG

FIG. 11

SEQ ID NO:9 TRP427-GLY431

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTTCGTTTCGCCGACGGCCGTGGAGAACGCTGTGGGTGACCGTGTACTACGGCGTGCCTGT
 TGGAAAGGCCACCAACCCACCTGTTCTGGCCAGCAGCAGCCACGGGAGCTACAGCACCGAGGT
 GCACAACTGTGGGCCACCCACGCTGCGTGCCTGCCACCGACCCCAACCCCAGGAGATCGTGT
 GGAGAACGCTGACCGAGAACCTCAACATGTGGAAGAACACATGTGGAGCAGATGCACGAG
 GACATCATCAGCTGTGGGACAGAGCCTGAAGGCGCTGCGTGAAGCTGACCCCCCTGTGGGT
 ACCCTGACTGCACCAACCTGAAAGAACGCCACCAACCCAAAGAGCAGCAACTGGAGAGAT
 GGACCGCGGAGATAAGAACCTGAGCTTCAAGGTGACCAAGCATCGCAAAAGATGC
 AGAAAGGAGTACCCCTGTCTAACAGGTGACGTGGTGCCTATCGACAACGACAACACCGC
 TACAAGCTGATCAACTGCAACACCGCTGATCACCCAGGGCTGCCCAAGGTGAGCTTGA
 GCCCCATCCCATCCACTACTGCGCCCGGCCGCTGCCATCTGAAGTGCACAGCAAAAGAA
 GTTCAACGGCGACGGCCCTGCAACACCGTGAAGCACCGTGCAGTGCACCCACGGCATCCGC
 CGCTGGTGAACCCAGCTGCTGCAACGGCGCTGGCGAGGAGGGCTGTGATCCGC
 AGCGAGAACCTCACCGACAACGCCAACGACCATCATCGTGCAGCTGAAGGAGAGCGTGGAGAT
 CAACCTGACCCGCCAACACAACCCGGCAAGGATCATCACCATGGGCCCCGGCGCT
 TCTACGGCACCGGCCACATCATCGGCCATCATCGGCCAGGGCACTGCAACATCAGCGCGAG
 AAGTGGAAACAAACCCCTGAAAGCAAGTGTGACCAAGCTGCAAGGGCCAGTGTGGCAACAGAC
 CATCGTGTCAAGCAAGCAGCGAGCGGCCAGCCCGAGATCGTGTGACAGCTTCAACTCG
 GCGGGCGAGITCTTCACTGCGCAACAGCACCCAGCTGTTCAACAGCACCTGAAACACCATCG
 GCCCCAACACCAACGGCAATCGGCCATCACCTGCGCTGGCCGATCAAGCAGACATCAACCGCT
 GGGGGCGGAAGCCATGTAGCGCCCCCATCGGCCAGATCGCTGCAAGCAGCAACATC
 ACCGGCCTGCTGCAAGGCCGCAAGGGAGATCAGCAACACCAACCGAGATCTTCCG
 CCCGGCGGCCGCGACATCGCGCAACTCGGCCAGCGACCTGTAACAGTACAAGTACAAGTGTGTA
 AGATCGAGCCCCCTGGCGTGGCCGCAACCTGCAAGGCCAAGGGCAAGCGCGCTGGTGTGCAAGGGAGAAG
 CGCGCGAGCTGACCTGAGCTGGCGTGGCCGCTGGCGCTGGCGAGCACCATGGGC
 GCGCCGAGCGCTGACCTGAGCTGGCGAGCGGCCAGTGTGAGGGCATCTGCAAGCAGCA
 GAACAACTGCTGCGGCCATCGAGGCCAAGCACACTGTCAGCTGACCGTGTGGGCA
 TCAAGCAGCTGCAAGGCCGCTGGCTGCTGGCGTGGAGGCTACTGAGGGACAGCTGCTG
 GGCATCTGGGCTGCAAGCTGATCTGCAACCGCGCTGGCTGGAGACCGCAAGCTG
 GAGCAAACAGAGCTGACCAAGATCTGAAACACATGACCTGAGATGGAGTGGAGCGCGAG
 ATCGACAACCTACACCAACTGATCTACACCCATGACGAGGAGAGGCCAGAACAGCAGGAGAA
 GAAAGCAGCAGGAGCTGCTGGAGCTGGACAAGTGGCCAGCCGTGTTAACTGGTGTGACATCA
 GCAAGTGGCTGTGGTACATCAAGATCTTCACTGATCTGTTGGGGGGCTGGGGCTGCGCA
 TCGTGTTCACCGTGTGAGCATCTGTAACCGCTGGCCAGGGCTACAGCCCCCTGAGCTTCA
 AGACCCGCTTCCCCCGCCCCCGGCCGACGGCCCGAGGGCATCGAGGGAGGAGGGCG
 GAGCGCGACGGCGACGGCGAGCAACCCCTGTGCAAGGCCCTGCTGGCCCTGATCTGGAGCA
 CCTGCGCAGCTGTGCTGTGAGCTACACCGCTGCGCACGCTGATCTGATCGCCGCCG
 CATCGTGTGGAGCTGTGGCCGCGCCGCGCTGGAGGGGGCTGGAGAATGACTGGGGCAACCTGCTGC
 AGTACTGATCAGGAGCTGAAAGAACAGCGCGTGTAGGCTGTGACGCGCATCGCACTGCC
 GTGGCGAGGGACCGACCGCATCGAGGTGGCCAGCGCATCGAGCGCCGCTTCTCGCA
 CATCCCCCGCCGACATCGCCAGGGCTGAGCGCGCCCTGCTGTAACTCGAG

FIG. 12

SEQ ID NO:10 ARG426-GLY431

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGAGCA
 GTCTTCTTTGCCCCAGGCCGCTGGAGAAGCTGTGGGTGACCGTGTACTACGGCGTGCCGTG
 TGGAAAGGAGGCCACACCCACCTGTTCTGGCCAGCGACGCCAACGCCCTACGACACCGAGGT
 GCACAACTGTGGGCCACCCACGCCCTGCGTGCACCCAGGCCAACCCCAAGGAGATCGTGT
 GGAGAACCTGACCCGAGAACACTTCAACATGTGGAAAGAACACATGTGGAGCAGATGCAACGAG
 GACATCACTACGCTGTGGGACAGAGCCTGAAGGCCCTGCGTGAAGCTGACCCCCCTGTGGTG
 ACCCTGACTGCAACCAACCTGAAGAACCGCCACCAACCCAAAGGAGCAACTGGAAAGGAGAT
 GGACCGGGCGAGATCAAGAACTGCACTGCAAGGTGACCAACGATCGCAACAAAGATGC
 AGAAAGGATACGCCCTGTCTACAAGCTGGAAGTGTGGGCCATCGACAAACGACAACACAGC
 TACAAGCTGATCAACTGCAACACCAAGCGTGTACCCAGGGCTGCCCCAAGGTGAGCTTGA
 GCCCATCCATCCATCTGCCCCCGCCGCTGCGCATCTGAAGTGCACAGCAAAAGAA
 GTTCAACGGCAGCGGCCCTGCAACCACTGTGAGACCGCTGGCAGTGCACCCACGGCATCCGCC
 CCTGTTGAGCACCCAGCTGCTGCTGACCGGCAGCTGGCGAGGGGGCTGGTGTATCCGC
 AGCGAGAACTTACCGACAAACGCCAACGACATCATCGTGCAGCTGAAGGAGAGGGTGGAGAT
 CAACTGCAACCCGCCAACAAACACCCGCAAGACATCACATCGGCCCCAGATCGCAACATCG
 TCTACGCCACCGGCCAACATCATCGGCCGACATCGGCCAGGCCCCATCGCAACATCGAGCG
 AAGTGGAAACAAACCCCTGAAGCAGATCGTGAACCGAGCTGCAAGGCCAGTTCGGCAACAGAC
 CATCGTGTCAAGCAGAGCAGCGGCCAGGGCCAGACCTGTGATGCAACAGCTTCAACTGCG
 GCGCGAGATTCTCTACTGCAACAGCACCCAGCTGTTCAACAGCACCTGGAAACACCATCG
 GCGCCAAACAAACCAACCGCAGCATCACCTGGCCCTGGCAGCATCAAGCAGATCATCAACCGC
 GCGCGCGCAAGCGCATGTCAGCGGCCCATCGGCCGAGATCGTGCAGCAGCAACAT
 CACCGGCCCTGTCTGACCCCGACGGCGCAAGGAGATCAAGCAACACCCAGGAGATCTTC
 GCGCCGGGGGGCGCAATCGGCCGACATCGGCCAGCTGGCAGCGAGCTGTCAGAACAGGTG
 AAGATCGAGCCCCCTGGCGCTGGGCCACCAAGCGCAACGGCGCGCTGGTGCAGCGCGAGAA
 GCGCGCCGACCTGGCTACCTTGGGCCCTGTGTCAGGGGCCCTGGGCCCGCAGCACCATGG
 CGCGCCGCAAGCTGACCTTGACCGTGCAGGCCGCAAGCTGTCAGGCGCATCGTGCAGCAGC
 AGAACACACTGTGCGGCCATCGAGGCCAGCGACGACCTGTGCACTGAGCGCTGTGGGGC
 ATCAAGCAGCTGCAAGCCCGCTGGCTGGAGCGCTGGCTACCTGAAGGAGCCAGCAGCTGCT
 GGGCATCTGGGGCTGCAAGGGCAAGCTGATCTGACCAACCGCCGTGCGCTGGAAACCGCAGT
 GGAGCAACAAGAGCTGGACCAATCTGGAACACATGACCTGGATGGAGTGGAGGGCGGA
 GATCGACAACACTACCAACCTGATCTACACCTGATCGAGGAGAGCCAGAACAGCAGGGAGA
 AGAACAGCAGGAGCTGCTGGAGCTGGACAAAGTGGGCCAGCCCTGTGGAACCTGGTGCACATC
 AGCAAGTGGCTGTGTCATCAAGATCTTCATCATGTCCTGGGCCCTGGTGGGCCCTGGC
 ATCGTGTTCACCTGCTGAGCATCGTGAACCGCCTGGCCAGGCTACAGCCCCCTGAGCTTC
 CAGACCCGCTTCCCGCCCCCGCGGCCGAGCGCCCGAGGGGATCGAGGAGGGGGCG
 CGAGCGCGACCCGACCCGAGCAAGCCCCCTGGTGCAGGCCCTGCTGGCCCTGATCTGGAGC
 ACCTGGCAGCGCTGTGCTGTCAGTACCAACCGCCGTGGCAGCTGATCTGATCGGCC
 GCATCGTGGAGCTGCTGGGCCCGCGGGCTGGAGGGCCCTGAAGTACTGGGCAACCTGCTG
 CAGTACTGGATTCAGGAGCTGAAAGAACACAGCCGTGAGCTGTCAGGAGTGGCCAGCG
 CGTGGCCGAGGGCACCGACCGCATCGAGGTGGCCAGCGCATCGGCCAGGCTCGTGAACCG
 ACATCCCCGCCGACATCGGCCAGGCTCGAGGGCTCGTGAACCGCAGCTGGCCGCGCTTC
 TGC

FIG. 13

SEQ ID NO:11 ARG426-GLY431B

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTCTTTCGCCAGGGCCGGTGGAGAAGCTGTTGGTGAACGGTACTACGGCGTGCCCGTG
 TGGAAAGGAGGCCACCAACCTCTGGCCAGCAGGCCAAGGCTACGACACCGAGGT
 GCACAACTGTGGGCCACCCACCGCTCGGTGCCACCGACCCCAACCCCAGGAGATCGTCT
 GGAGAACGTGACCCAGAAACTTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCAAGAG
 GACATCATCAGCTGTGGACAGAGCCTGAAGGCCCTCGTGAAGCTGACCCCCCTGTGCGTG
 ACCCTGCACTGACCAAACTTGAAGAACGCCCCAACCAAGAGCAGCAACTGGAGAGGAT
 GGACCGCGGGAGATCAAGAACGTGAGCTTCAAGGTGACCAACAGGATCGGCAAAAGATGC
 AGAAGGAGTACGCCCTGTCTACAAGCTGGACGTGGTGCCTACGACAAACGACAAACACAGC
 TACAAGCTGATCAACTGACACCAACAGGCTGATCACCCAGGGCTGCCCAAAGGTGAGCTTGA
 GCCCCATCCCCATCCACTACTGCCCGCCCGCCTCGGCATCTGAATGTCACAGCAAGCAAAGAA
 GTTCAACCGGCGGGCCCTGACCAACGTGAGCACCCGCTGAGTGCACCCACGGCATCGCC
 CGCTGGTGAACGGCAGCAGCTGCTGTAACGGCAGCAGCTGGCCAGGGCGTGGTGAATCGC
 AGCGAGAACTTCAACCGAACAGCAGGACATCATCTGTCAGCTGAAGGAAGGCTGGAGAT
 CAACTGACCCGGCCCCAACAAACACGGCAAGACGACATCACCATGGGCCCCGGGGCGCT
 TCTACGGCACCGGGACATCATCGGCAGATCCCGCAGGCTCGACACATCAGCGCAG
 AAGTGGAAACACCCCTGAAGCAGATCGTGAACAGCTGCAGGCCAGTTGGCAACAAGAC
 CATCGTGTGAGCAGAGCAGCGCCGGGAGGAGATCTGATGACAGCTGACACTCG
 GCGGGAGTTCTACTCGAACACGACCCAGCTGTCACAGCACCTGGAAACAACACCATCG
 GCCCCAACAAACCAACGGCACATCACCTGGCCCTGGCCGACATCAAGCAGATCATCAAACCG
 GGCAGCGCCAGGCGAATGACGGCCCCCATTCGGCCAGATCGTGCAGCAAGAACAT
 CACCGGCTGCTGACCCGGACGGCGGCAAGAGGAGTCAACGCAACACCCAGGAATCTTC
 GCCCCGGGGCGCGCATCGCGACAACTGGCGCAGCGAGCTGACAAAGTACAAGGTGIG
 AAGATCGAGGCCCCCTGGCGTGGGCCCCAACAAAGGCCAACGGCCGGTGGTGCAGGGAGAA
 CGCGCGCTGAGACCTGGCCCTGGCGCATGGTCTGGGCTCTGGCGCCGCCGGAGACCATGG
 CGGGCGCAGCTGACCTGACCGTGCAGGGCCCGAGCTGCTGAACGGGCGATCGACAGC
 AGAACAACTCTGCTGGCGCCATCGAGGCCAGCAGCACCTGCTGAGCTGACCGTGTGGGG
 ATCAAGCAGCTGACGGCCCGCTGCTGGCCCTGGAGCGCTACCTGAAAGGACCAAGCAGCTG
 GGGCATCTGGGGCTCGACGGCGAACAGCTGATCTGACACCGCCGTGCCCCGGTGAACGG
 GGAGCAACAGAGCTGACCCAGATCTGAAACAATGACCTGATGGAGTGGAGGAGCGGA
 GATCGACAACACTACCAACCTGATCTACACCTGATCGAGGGAGAGCCAGAACAGCAGGAGA
 AGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCTGTGGAACCTGGTGCACATC
 AGCAAGTGGCTGTTGATCTCAAGATCTTCATCTGATCTGTTGGGGGCGCTGGTGGCG
 ATCTGTTGACCTGGCTGAGCATCGTGAACGGCGCTGGCCGAGGGCTACAGCCCCCTGAGCTT
 CAGACCCGCTTCCCCGCCCCCGGGCCCCGACGGCCCCGAGGGCATCGAGGGAGGAGCG
 CGAGCGCGAGCCCGAGCCGACAGCCCCCTGGTGCACGGCCCTGCTGGCCCTGATCTGGAG
 ACCTGCGCAACCTGTCCTGCTGTCAGTACACCGCCCTGCGCGACCTGATCTGACCGGCC
 GCATCTGAGGCTGCTGGAGGCGCTGGTGGGGCGCGCTGGAGGGCCCTGAAGTACTGGGG
 CAGTACCTGATGCCAGGAGCTGAAAGAACAGCGCCGTGAGCTGTTGACGCGCATCGGC
 CGTGGCCGAGGGCACCAGCCGACATCGAGGGTGGCCAGCGCATCGGCCGCGCCCTG
 ACATCCCCCGCCGATCCGCCAGGGCTGAGCGCGCCCTGCTGTAACCTCGAG

FIG. 14

SEQ ID NO:12 ARG426-LYS432

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTCGCTGTGGAGCA
 GTCTTCGTTGCCAGCGCCGTGGAGAAGCTGTGGTACCGTGACTACGGCTGCCGTG
 TGGAAAGGAGGCCA CCA CCA CCGCTGTCTGCGCCAGCGA CGCCA AGGCCTA CGA CACCGAGGT
 GCA CAA CGTG TGGGCCA CCA CGCTGCGTGGCCACCGA CCCCACCCCA CGGAGATCGTGCT
 GGAGAACGTGACCGA GAACCTTCACATGTGGAAGAACACATGGTGGAGCA GTGACAGAG
 GACATCATCAGCCTGTGGAGCAAGGCTGAAGGCTGCGTGAAGCTGACCCCTGTGCGT
 ACCCTGCACTGCA CCAACCTGAAAGGAGGCCA CCAACCAAGAGCAGAACCTGGAAGAGAG
 GGACCGGGCGA GATCAAGGAACCTGCAAGGTGACCCAGCATCGCAACAAAGATGC
 AGAGGGATCA CGCTGTGGTCTCA CAA CGTGAGCGTGGTGGCCATCGACAACGACAACACAGC
 TACAAGCTGATCAACTGCAACACACAGCGTGATCACCCAGGCTGCCCAAGGTGAGCTTCGA
 GCCCCTCCCTTCAACTCTGCGCCCGCCCGCTCCGCAAGCTGCAAGTGCACAGCAAGAA
 GTTCA CCGCAGCGGCCCTGCACCAAGCTGACCGTGCAGTGACCCACCGCATCGCC
 CGTGTGAGCAACCCAGTGTGCAACGGCAGGCTGGCGAGGAGGGCGTGTGATCGC
 AGOGAGAACCTTCACCGAACACGCCAAGGACATCTGTCAGCTGAGCTGAAGGGAGAGCGTGGAG
 CAACTGCA CCGCCCAAACAAACACCCGCAAGAGCATCACATCGTGGCCCGCCCGCCT
 TCTACCGCACCGGCGACATCGCGCCAGGCCCAGCTGCAACATCAGCGCCGAG
 AAGTGGAA CAAACCTGAAAGCAGATCGTGGACAAAGCTGCAAGGCCAGTTCGGCAACAAAGAC
 CATCGTGTCAAGCAGAGCAGCGGGCGGAGACCCGAGATCGTGATGCAAGCGTCAACTCGC
 GCGCAGAGTCTCTCACTGCAACAGCAGGCCAGCTGTCACAGCACCTGGAAACAAACACATCG
 GCCCCAA CAA CACCAA CGGCCACCATACCGCCCTGCGCATCAAGCAGATCATCAACCGC
 GCGGCGAACAGGCCATGTCAGCCCGGACGGCCAGATCGCTGCAAGCAGAACAT
 CACCGCCCTGCTGTGACCCGGACGGCGCAAGGAGATCGACAACACCA CGGAGATCTTC
 GCCCCGGCGGGCGACATCGCGACA CTGGCGCAGGGAGCTGATCAAGTACAAGGGTGT
 AAAGATCGAGCCCCCTGGCGTGAAGGCCCAACCGAAGCGCAGCGCGCTGGAGCGCGGAGA
 GCGCGCGTGA CGCCCTGGCGCATCTGCTCTGGCGCTCTGGCGCCCGCGCAGCACCATGG
 CGCCCGCAGCGCTGACCCGTGACCGTGCAGGCCAGCAGCTGCTGCAAGCTGACAGCAGC
 AGAACAACTCTGCTGCGCCATCGAGGCCAGCAGCACCTGCTGCAAGCTGACCGTGTGGGC
 ATCAAGCAGCTGCAAGGCCCGCTGCTGGCTGTGAGGCCATCTGAAAGGACAGCAGCTG
 GGGCATCTGGGCTGCAAGGGCAAGCTGATCTGCAACACCCCGTGGCTGAAACCCGAGCT
 GGAGCAACAGGCCAGCATCTGGAAACAACTGACCTGGATGGAGCTGGAGCGCGGA
 GATCGACA ACTAACCAAACCTGATCTACACCTGATCGAGGAAGGCCAGAACAGCAGGAGA
 AGAACGAGCAGGAGGAGCTGCTGGACAACTGGGCCAGCTGTGGAGCTGGAACTGGTTCAGACATC
 AGCAAGTGGCTGTGGTACATCAAGATCTCATGATCGTGGCCGCTGTGGCCCTGCGC
 ATCGTGTTCACCGTGTGAGCATCGTAACCCGCTGCGGCCAGGGCTACAGGCCCTGAGCTTC
 CAGACCCGCTCCCGCCCCCGGCCGGCAGCCGAGGCCATCGAGGGAGGGCG
 CGAGCGCAGCCGCGACCGCAGCACGCCCTGGTCACGCCCTGCTGGCCCTGATCTGGACG
 ACCCTGGCGACGCTGTGCTGTGCTGCTGCTGACCTACCAACGCCCTGGCGACCTGATCTGACCGGCC
 GCACTGTTGGAGCTGTGCGCCGCCGGCTGGAGGGCCCTGAAGTACTGGGGCAACCTGCTG
 CAGTACTGGATCAGGAGCTGAAAGAACAGGCCGTGAGCCTGTTGACGCCATCGCCATGCG
 CGTGGCGAGGGCACCGCATCGAGGTGGCCAGGCCATCGGCGGCCCTGCTGAACTCGAG
 ACATCCCCCGCCGATCGCCAGGGCTTGAGCGCCGCTGCTGTAACCTGAG

FIG. 15

SEQ ID NO:13 ASN425-LYS432

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTTGCTGCTGTGGAGCA
 GTCTTCGTTGCCCAAGCGCGTGGAGAAGCTGTTGGTGTACCGCTGACTACGGCTGCCGTG
 TGGAAAGGAGGCCACCAACCCACCTGTTCTGCGCCAGGACGCCAACGGCTAAGCACCGAGGT
 GCACAACGTTGGGCCACCCACGCCCTGCGTGCACCGACCCCAACCCCAGGAGATCTGCTG
 GGAGAACGTTGACCGAAGAACCTTCAACATGTGGAAGAACACATGTGAGCAGATGCAAGAG
 GACATCATCAGCTCTGTTGGAGCCAAGGCTGACGCCAGGCTGTGAAGCTGACCCCCCTGCTG
 ACCCTGCACTGACCAACCTGAAAGACGCCAACACAACAGAGCAGCAACTGGAAGAGGAT
 GGACCGCCGGAAGATAAGAACCTGCACTTCAGGTAAGGTGACCCAGCAGATCCGCAAACAGATG
 AGAACGGATGACCGCTGTCTCAACAGTGTGAGCTGTTGCCCATCGACAAACGACAACACCAGC
 TACAAGCTGATCAACTGAAACACAGCTGATCACCCAGGCTGCCCAAGGTGACCTGCA
 GCCCCATCCCCATCAACTGCGCCGCCGGCCGCGGCCATCTGAAGTGTGAAACGCCAACAGAA
 GTTCAACGGCAGGGGCCCTGCACCAACAGTGTGAGCACCGTGCAGTGCACCCACGGCATCCGC
 CGTGTGAGCACCAAGCTGCTGTAACCGCAGCTGGCCAGAGGAGGGCTGTTGATCCG
 AGCGAGAACCTAACCGAACAGGCCAACGATCATCTGTCAGCTGAAGAGGAGGGCTG
 CAACATGCAACCCGCCAACACAACACCCCGCAAGGACATCACCATGCCCGGCCGGCT
 TCTACGCCACCGCGACATCGGCCGACATCGGCCGAGGCCACTGCAACACATCAGGGCAG
 AAGTGGAAACACCTGTAAGCACGATCTGTAACAGCTGCAAGGCCAGTTCGGCAACAGAC
 CATCGTGTCAAGCAGAGCAGCGGGCGGAGCCCGAGATCTGATGCAAGCTTCAACTGCG
 GCGGGAGTTCTCTCACTGCAACAGCAGGCCAGCTGTTCAACAGCACCTGGAAACACCAATCTG
 GCCCCAACACAACACCGCACCATCACCTGCCCTGCCGCATCAAGCAGATCATCACGCC
 CCAACCGCATGTAACGGCCCCCATCGGCCGGAATCGCTGCAAGCAGAACATCACCGCC
 TGTGCTGACCCGGCACGGCAAGAGATCGAACACACCCAGGATCTTCCGGCCGG
 GGGCGCGACATGCCGACAACATCGCGCAGCAGGCTGTAACAGTACAAGGTGTAAGATCGA
 GCCCCCTGGGCTGGGCCCCCAAGGCCAACAGGCCAGCGCCGCTGGTGCAGCGGAGAACGGCCG
 TGACCTTGGGGCCATGTCTGGCTTCCTGGGCCCGGCCAGCTGCTGAGGGCATCTGCA
 GCGTGCACCTGACCGTGCAGGCCAGCAGCACCTGTCAGCTGAGCGAACAC
 CTGCTGGCGCATCGAGGCCAGCAGGCCAGCTGCTGGAGCAGTACTGAAAGGACAGCAGTGTG
 GCTGCAACCGCCGTGCTGGCGTGGAGCAGTACTGAAAGGACAGCAGTGTG
 GGGGCTGAGGGCAAGGGATCTGACCCACGCCGCTGGCTGGGAAACGCCAGCTGAGC
 AACAGGCTTGGGACAGATCTGGAAACATGACCTGATGGTGTGAGTGGGAGGCCAGATGACAA
 CTACACCAACCTGATCTACACCTGATCGAGGGAGGCCAGAACAGCAGGAGAACAGCAG
 AGGGAGCTGCTGGAGCTGGACAAAGTGGCCAGCTGTTGAACTGGTTCGACATCGAACAGTGG
 CTGTTGATCATCAAGATCTCATGATGCTGGGGCCCTGTTGGGCTGCGCATCTGTTG
 ACCGTGCTGAGCATGTAACCGCTGCGCCAGGGCTACAGCCCCCTGAGCTTCAAGGCC
 TTCCCCCCCCCCCCGGGCCAGGCCAGGGCTGAGGGAGGCCAGGCCAG
 CGCGCACCGCACAGGCCCTGTTGACGCCCTGCTGCCCTGATCTGGGAGCAGCTGCCAG
 CCTGCTGGCCGGCGGCCGGCTGGAGGCCCTGAAAGTACTGGGGCAACCTGCTGAGTACTGG
 TCCAGGGAGCTGAAGAACAGGCCGTGAGGCTGTTGACGCCATGCCCATCGCCGTGGCGAG
 GGACCGACCGCATCTGAGGTGGCCAGGCCATCGGGCGGCCCTGCTGACATCCCCCGC
 CGCATCGGCCAGGGCTTGAGGGCGGCCCTGCTGTAACCTGAG

FIG. 16

SEQ ID NO:14 ILE424-ALA433

GAATTCGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTTCGTTGCCACGGCGCTGGAGAAGCTGTGGGTGACCGTGTACTACGGCTGCCGTG
 TGGAAAGGAGGCCACCAACCAACCCCTGTTCTGGGCCAGCGACGCCAACGGCTAACGACACGGAGGT
 GCACAACGTGTGGGCCACCCACGCCCTGCCGTGCCAACCGACCCCAACCCCAGGAGATCGTGTCT
 GGAGAACGTGACCGAGAACCTCAACATGTGGAAGAACACATGGTGGAGCAGATGCAGAG
 GACATCATCAGCCTGTTGGACCCAGAGCCTGCTGAAGGCTCTGCTGAAGCTGACCCCCCTGTGGTG
 ACCCTGACTGACCAACCTGAAAGAACGCCAACACCCAAGAGCAGCAACTGGAGAGAT
 GGACCGCCGCGAGATAAGAACCTGCAAGGTGACCCAGCATCGCAACAGATGC
 AGAAGGAGTACCGCCCTGTCTACAAAGCTGGACCTGGTGCCTCATGACAACGGACAACACAGAC
 TACAAGCTGATCACTGAAACACCCAGCTGATCACCCAGGCTGCCCAAGGTGAGCTCGA
 GCCCCATCCCCATCCACTACTGCGCCCCCCCAGCTGCCATCTGAAAGTGTCAACAGCAAGAA
 GTTCAACGGCAGGCCGCCCTGCCACCAACGTGAGCACCTGCTGAGTGCACCCACGCCATCGGCC
 CCGTGGTGAACCCAGCTGCTGAACGGCAGCTGGCGAGGAGGGGTGTTGATCGC
 AGCGAGAACCTACCGAACAGGCCAACATCGTGCAGCTGAAGGGAGAGCTGGAGAT
 CAACTGACCCGCCAACACACACCCGCCAACAGGATCATCGTGCAGCTGAAGGGAGAGCTGGAGAT
 TCTACGGCACCGCGACATCAGGGAGCATGCCAGGCCACTGCAACATCAGCGCGAG
 AAGTGGAAACAACACCTGAAAGCAGATCGTGAACAGCTGAGGCCAGGTTGCCAACAAAGAC
 CATCGTGTCAAGCAGAGCAGCGGCCGAGCCAGATCGTGTGACAGCTTCAACTGCG
 GCGGGAGGTTCTCTACTGCAACAGCAGCCAGCTGTCACAGCACCTGGAAACAACACCATCG
 GCCCCAACACACCAACGGCACCATCCCTGCCGCCATCAAGCAGATCATCGGCC
 GCCATGATCGCCCCCCCCCATCGGCCGAGATCGTGCAGCAGAACATCAGCGCTG
 CTGACCCGGCACGGCGAAGAGATCAGCAACACCCAGGATCTCCGCCCGGGCG
 CGACATCGGCACAACTGGCGACCGAGCTGATCAGTACAAGGTGTTGAAGATCGAGCCC
 TGGGGTGGCCCCCACCAGGGCAAGGGCCAGGGCTGTGTCAGGGAGAAGGGCGCGGTGACC
 CTGGCGCCCATCTCTGCTCTGGGCTTCTGGGCCGCCGCCAGCACCATGGGCCCGACGCTG
 ACCCTGACCTGTGAGGCCGCCAGCTGCTGAAGGGCATCTGCAAGCAGAACAACTCTG
 GCGCGCATCGAGGGCCAGCAGACCTGCTGAGCTGACCGTGGCTGGGGCATCAAGCAGCTG
 AGGGCCGGTGTGGCGCTGGAGCAGCTGAAGGGACCAAGCAGCTGCTGGCATCTGGGG
 TGCAAGCGGAAGCTGATCTGCAACCCCGCGTGGCTGGGAACGCCAGCTGGAGCAACAAAGAG
 CCTGGACCCAGATCTGAAACATGACCTGGATGAGTGGGGAGCGCGAGATCGACAACTACA
 CCAACCTGATCTACACCTGTGAGGAGAGCCAGAACCCAGCAGGAGAAAGACGAGCAGGA
 GCTGCTGGAGCTGCAAGTGGCCAGCTGTGGAAGCTGGTTCGACATCGAAGATGGCTG
 GTGACATCAAGATCTCATGATCGTGGCGCGCTGGGCTCTGGCATCTGTTGTCACCG
 TGCTGAGCATCTGAAACCCCGTGGCCAGGGCTACAGCCCCCTGAGCTTCCAGACCCCGCTCC
 CGCCCCCCCAGGGCCGCCAGCCCCGAGGGCATCGAGGGAGGGGGAGCGACCTGCGCAGCGCTG
 TGCTGTCTAGCTTACCAACGGCTGGCGACCTGTGATCTGACGCCGCCAGCTGTGGAGCTG
 CTGGGCCGCCGGCTGGGAGGGCCCTGAAGTACTGGGCAACCTGCTGAGTACTGGATCCA
 GGAGCTGAAGAACAGGGCCGTGAGCTGTTGACGCCATGCCATGCCGTGGCCAGGGCA
 CCGACCGCATCATCGAGGTGGCCAGGGCATCGGCCGCCCTCTGACATCCCCCGCGA
 TCCGGCCAGGGCTTGTGAGCGGCCCTGCTGTAACTCAG

FIG. 17

SEQ ID NO:15 ILE423-MET434

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGCTGTGGAGCA
 GTCTTCGTTGCCAGGCCGTTGGAAAGCTGTTGGGTGACCGTGTACTACGGGTGCCGTG
 TGGAAAGGAGGCCACCAACCCCTGTTCTGCCAGCGGAGCCAAAGGCTTACGACCCGAGGT
 GCACAACGCTGTGGGCCACCCACCGCTGGTCCACCGGACCCCAACCCCGAGGAGATCGTGT
 GGAGAACCTGACCGAGAACCTCAACATGTGAAAGAACACATGTGAGCAGATGACAGAG
 GACATCATCAGCTGTGGGACCCAGAGCTGAAGGCTCTGGGTGAAGCTGACCCCTGTGGTG
 ACCCTGCACTGACCAAACCTGAAAGAACCCACCAACACCAAGAGCAGCAAATGGAAGGAGAT
 GGACCGCGGGAGATCAAGAACCTGCAATGAAGTGGACCCAGCATCCGCAACAAAGATGC
 AGAAAGGAGTACGCCCTGTTCTACAGCTGGACTGTGCCCCATCGACAACGACAACACCGC
 TACAAGCTGATCACTGCAACACCCAGCGTGTACCCAGGCCCAAGGTGAGCTTG
 GCCCCATCCCCATCACTACTGCGCCCGCCCGGCTTCGCCATCTGAAAGTGAACACGACAAGAA
 GTTCAACGGCAGGGCCGCCCTGCAACCTGAGCACCTGGTGAAGCTGACCCACGGCATCGGC
 CGTGGTGAAGCACCCAGCTGCTGTGAACGGCAGCTGGCGAGGAGGGCTGTTGATCCGC
 AGCGAGAACCTACCGCACCGAACGACCATCTGCACTGAGCTGAAGGGAGAGCTGGAT
 CAACTGCAACCCCGGAAACAAACACCCGCAAGAGCATACTGACCCGGCCGGCGCG
 TCTACGGCACCCGGGACATCGGCCACATCGGCCACGCCCCACTGCAACATCAGCGGAG
 AAGTGGAAACACCCCTGAAGCAGATCGTGAACAGCTGGCAGGGCCAGITCGGCAACAGAC
 CATCGTGTCAAGCAGAGCAGCGGGGGACCCCGAGATCGTGTATGACAGCTTCAACTCG
 CGGGGAGACTTCTACTGCAAGCACAGCCAGCTGGTCAAGCAGCTGGAACCAACACCTG
 GCCCCAACAAACACCAACGGCACCATACCCCTGCCGCGCATCAAGCAGATCGGCCGATG
 TAGCGCACCCCGGATCGGCCGAGATCGCTGCAAGCACAATCACCGGCCCTGCTGCTGACC
 CGCGAGCCGGGCAAGGGAGATCGAACACCCAGAGATCTCCGCCCCGGGCCGGGACAT
 CGCGCACAACTGGCGAGCAGACTGTAACAGTACAAGGTGGTGAAGATGAGCTGGCCCCCTGGCG
 TGGCGGCCCCAACAAAGGGCGGCGCTGGTCAAGGGAGAAGCGCGCCGTGACCTGGC
 GCGCATGTTCTGGGCTTCTGGGCCGCGGCCGAGCACCATGGGCCCGCAGCTGACCTG
 ACCGTGCAAGGCCGAGCAGCACCTGCTGCACTGAGCGGATCGTGCAGCAGCGAACAACTG
 CATCGAGGCCGAGCAGCACCTGCTGCACTGAGCGGATCGTGTGGGGCATCAAGCAGCTGCA
 CGCGTGTGGCGTGGAGCGTACCTGAAGGGACCGAGCAGCTGGCATCTGGGCTCGAC
 GCGAAGCTGATCTGCAACCCGGCTGCCCCCTGGAAAGCGGAGCTGGAGCAAAAGGCTGGA
 CCAGATCTGAAACACATGACCTGGATGGAGCTGGAGCGCGAGATGACAACATACCAA
 TGATCTACACCCCTGATCGAGGGAGAGCCAGAACCCAGCAGGAGAAGAACAGAGCAGGAGCTG
 GAGCTGGAAAGTGGGCCAGGCTTGTGGAACTGTTCTGACATCGACAAGTGGCTGGTACAT
 CAAGATCTTCACTGATCGTGGGCCCTGTTGGGCTGCGCATGTTTACCGTGTGAG
 CATCGTGAACCCGGTGGCCAGGGCTACAGCCCCCTGAGCTTCAAGCCCGCTTCCCCGCC
 CGCGGCCCCGAGCCGGGAGGGCATCGAGGGAGGGGGCGGAGCGCGAACCGGACCG
 AGCAGCCCCCTGGTGCACGGCTGCTGGCCCTGATCTGGGACGACCTGGCAGCGCTG
 TTCACTGACCCAGCGCTGGCGGACCTGATCTGATTCGCGGCCGATCTGGAGCTGGCTGGC
 CGCGCGGCTGGAGGGGGCTGAAGTACTGGGCCACCCCTGCTGAGTACTGGATCCAGGAGCT
 GAAGAACAGGGCGTGAACCTGTTGACGCCATCGCCATCGCGTGGCCAGGGGACCGACC
 GCATCATCGAGGTGGGCCAGCGCATCGCCGCGGCCCTCTGACATCCCCCGCGCATCG
 AGGGCTTCACTGAGGCCGCGCTGTAACCTGAG

FIG. 18

SEQ ID NO:16 GLN422-TYR435

GAATTCGCCACCATGGTCAATGAAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTCTGTTGCCAGGCCGTGGAGAAGCTGGGGTGACGGTGTACTACGGCGTGCCTGCT
 TGGAAAGGAGGCCACCACCCCTGTTCTGGGCCAGCGAGCCAAAGGCCTACGACACCGAGGT
 GCACAAACGTTGGGCCACCACCGCTGGTGCCTACCGAACCCCAACCCCCAGGATCGTGT
 GGAGAACGTTGACCGAGAACCTAACATGGAAAGAACACATGGTGGAGCAGATGACAG
 GACATCATCAGCTGTGGGACCAAGGGCTGAAGGCTCTGGTGAAGCTGACCCCCCTGTGGT
 ACCCTGACTGACCAACCTGAAGAACGCCAACACCAAGAGCAGCAACTGGAAGAGAT
 GGACCCGGGGAGATCAAGAACGCTTCAAGTGGACCCAGCATCCGCAAAAGATGC
 AGAAAGAGTACCCCTGTTCTAACAGTGGACGTGGTGCCTACGACAAAGCACACCCAGC
 TACAAAGCTGATCAACTGCAACACCCAGCGTGTACCCAGGCCAAGGTGAGCTTGA
 GCCCCATCCCCATCTAACATGCGCCCGCCCGCTTCGACATCTGAAGCTGCAACGAAAGAA
 GTTCAACGGCACGGCCCCCTGACCAACGTTGAGCACCTGGTGCAGTGACCCACGGCATCGCC
 CGTGGTGAACCCAGCTGCTGTAACGGCAGCTGGCGAGGAGGGCTGTTGATCCGC
 AGCAGAACCTTACCCGAAACGCAACGACCATCTGTCAGCTGAAGGGAGGCTGGAGAT
 CAACTGCAACCCGCCAACAACACACCCGCCAACGAGCATCACCATGGCCCCGGCCGCT
 TCTACGGCACGGCGACATCGGCCATCGGCCAGGGCCACTGCAACATCGGCCAG
 AAGTGGAAACACCCCTGAAGCAGATCGTGACCAAGCTGCAAGGGCCAGTGTGGCAAAAGAC
 CATCGTGTCAAGGAGCAGCGCCGGCGAGCCCGAGATCTGAGTGCACAGCTTCAACTCGG
 GGGCGAGTTCTCTACTGCAACGACCCAGCTGTCACAGCACCCTGGAAACAAACCATCG
 GCCCCAACACACCCGCCAACGACCGAACATCGGCCATCGGCCAGCTGGCAACAGGGCGCTACGCC
 CCCCCCATCCGGGGCACATCGCAGCAGCAACATCACCGGCTGTGTCACCCGGAC
 GGGCGCAAGGAGTCAAGCACCAACCGAGATCTCCGGGGGGCGGCCGACATGGCGA
 CAACTGGCGCACGGAGCTGTAACAGTACAAGGGTGTGAAGATGAGGCCCCCTGGGGCTGGCC
 CCCCCAACACGGCAAGGGCGCGCTGGTGAAGCAGGAGAAGCAGCGCCGTGACCTGGGGCATGG
 TTCTGGGCTTCCTGGGCCCGCCGGCAGCATGGCGCCAGCATGGCGCCGAGCTGTGACCCCTG
 CAGGGCCGCAAGCTGTCAGCGGCATCGTGAGCAGCAACTGCAAGCAGCTGCAAGGGGGCG
 GGCCCCAGCAGCACCTGTCAGCTGACGGTGTGGGGCATCAAGCAGCTGCAAGGGGGCG
 TGGCGGTGGAGGGCTACCTGAAGGACAGCAGCTGCTGGGCACTCTGGGGCTGAGCGGCAAG
 CTGATCTGCAACACCGCCGTGCCCCCTGGAACGCCAGTGGAGCACAAGAGGCTGGAGCCAGAT
 CTGGAAACAACTGATCGTGGATGAGTGGAGCGCGAGATCGACAACACTACCAACCTGATCT
 ACACCCCTGATCGAGGAGAGCCAGAACAGCAGGGAGAGAACGAGCAGGGAGCTGCTGGAGCT
 GGACAAGTGGGGCAAGCTGTGGAACTGGTTCTGACATCGCAAGTGGCTGTGGTACACCAAGA
 TCTTCATCATGATCGTGGCGCCCTGGTGGGCTGCGCATCGTGTACCCCTGCTGAGCATCG
 TGAACGGCGTGGCCA GGCTCAAGCCCCCTGAGCTTCCAGACCCGGCTTCCCCGGCCCCCG
 GCCCCGACGCCCGAGGGCATCGAGGAAGGGGGCGAGCGCGAACCGGACCGCAG
 CCCCCCTGGTGCACGGCGCTGCTGGCCCTGATCTGGAGCACCTGGCAGGCTGTGGCTGTCAG
 CTACCAACGGCTGGCGCACCTGATCTCTGATCGCCGGCCGACATCGTGGAGCTGCTGGGGCG
 CGGCTGGGGAGGGCCCTGAAGTACTGGGCAACCTGCGAGTACTGTTGATCCAGGAGCTGAAGA
 ACAGGGCGTGAAGCTGTGAGGCCATGCCATCGCGTGGCCAGGGCACCGACCGCATC
 ATCGAGGGTGGGCCAGCGCATCGGCCGCGCTTCTGACATCCCCCGCGCATCGGCCAGGGC

FIG. 19

SEQ ID NO:17 GLN422-TYR435B

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTTCGTTTCGCCAGCGCCGTGGAGAAGCTGTGGGTGACCGTGTACTACGGCTGCCGTG
 TGGAGAGGAGGCCACCCACCCCTGTTCTGCCGACGCCAGGCCAACGGGATACGACCCAGGAGT
 GCACAAACGCTGTGGGCCACCCACGCTGCGTGCACCCAGGCCAACCCCCAGGGAGATCGTGTCT
 GGAGAACCTGACCGAGAACCTCAACATGTTGAAAGAACACATGGTGGAGCAGATGCCAGAG
 GACATCATCGCCTGTGGGACAGAGCCTGAAGGCTGCGTGAAGCTGACCCCCCTGTGCGTG
 ACCCTGACTGCAACCAACCTGAAAGACGCCACCAACCCAAAGAGCAGCAACTGTGGAGAGAT
 GGACGCCGGCAGATCAAGAACCTGCGACTTCAAGGTGACCAACAGATCCGCAACAGATG
 AGAAAGGATACGCCCTGTCTCACAACTGGTGGACCTGACAAACGACAACACCGC
 TACAAGCTGATCACTGCAACACCCAGCGTGTACCCAGGCCAACGGCTGCCAACGGTGA
 GCCCCATCCCCATCCACTACTGCCCGGCCGCGCTTGCCATCTGAAGTGCACAGCAAGAA
 GTTCAACGGCAGCGGCCCTGCAACACCTGAGACGCCCTGCGACTGACCCACGCCATCCGCC
 CCGTGGTGACGCCAGCTGCTGCTGAACCGCAGCTGGCGAGAGGGGTGTTGATCCGC
 AGCGAGAACCTCACCGACAACGCCAACACCATCTCGTGCAGCTGAAGGGAGAGCGTGGAGAT
 CAACTGACCCGCCCAACAAACACCCGCAAGGACATCACATGCCGACATCCCCGGCCGGCCT
 TCTACGCCACCGGCCACATCGCCGACATCGCCGAGCTGCAACATCAGCGCGAG
 AAGTGGAAACACCCCTGAAGGAGATCGTGAACAGCTGCAAGGGCAGATTCGGCAACAGAC
 CATCGTGTCAAGCAGAGCAGCGCCGAGCCCGAGATCGTGTACGCTCAACTCG
 GCGCCGAGATCTTCACTGCAACAGCACCCAGCTGTTCAACAGCACCTGGAACACACCATCG
 GCCCCAACACCAACGCCACCATCCCTGCCCTGCGCAGTCAAGCAGGGCCCCCTACGCC
 CCCCCATCCGGCCGAGATCTCGTGCAGCACCAACTCACCGCCGCTGCTGTCGACCCGG
 GCGCGAACAGGATCAGCAACACCCGAGATCTCGCCCGGCCGCGGAGATCGCCGAC
 AACTGGCCAGGGAGCTGTACAAGTACAAGGTGGAAGATCGAGCCCCCTGGGGCTGGCC
 CACCAAAGCCAAAGCGCCGGTGTGCAAGCGCGAGAGCGCGCCGCTGACCTGGGGCATGT
 TCTCTGGCTCTGGGCCGGCCGACCATGGCCGAGCTGGGCCCTGCGCAGCTGACCTGGCTGC
 AGGCCCGCCAGCTGCTGAGCGCACATCGTGCAGCACAGAACACCTGCTGCGCAGTCAG
 GCCCAGCAGCACCTGCTGAGCTGACCGTGTGGGCATCAAGCAGCTGCAAGGCCGGTGT
 GGCGCTGGAGGGCTACCTGAAGGACCAAGCAGCTGCTGGGCATCTGGGGCTGCAAGGGCAAGC
 TGATCTGCACCCACGGCGTGGCTGCAAGCCAGCTGAGCAACAAAGAGCTGACACAGAT
 TGGAAACAAACATGACCTGGATGGAGTGGAGGAGCCGAGATCGAAACTACACCAACCTGATCTA
 CACCTGATCGAGGAGAGCCAGAACAGCAGGAGAAAGACGAGCAGAGCTGCTGGAGCTG
 GACAAGTGGGCCAGCTGTGGAAGTGGTTCGACATCGAAGTGGCTGTTGACATCAAGAT
 CTTCATCATGATCGTGGGCCGCTGTGGGCCCTGCGCAGCTGTTCACCGTGTGACGATCTG
 GAACAGCGCTGCCAGGGCTACAGCCCCCTGAGCTTCCAGAACCCCTTCCCCGCCGCCGG
 CCCCCGACCCGCCAGGGCATCGAGGAGGAAGGGCGGCCAGCGCAGCCGAGCAGC
 CCCCCGGTGCACGGCTGCTGCCCTGATCTGGAGCACCTGCGCAGCTGCTGCTGCTG
 TACCAACGCCCTGCCGAGCTGATCTGATCGCCGCCATCGTGGAGCTGCTGGGGCCGC
 GGCTGGAGGGCCCTGAAGTACTGGGGCAACCTGCTGAGTACTGATCTGGAGGAGCTGAAGAA
 CAGCCCGTGAACGGCTGTCGACGCCATCGCCATCGCCGTCGCGCTGCCAGGGCACCGACCC
 CGAGGTGGCCAGGCATCGGCCGCCCTTCTGCAATCCCCGCCGATCCGCCAGGGCTT
 CGAGCGGCCCTGCTGAACTCGAG

FIG. 20

SEQ ID NO:18: LEU122-SER199; ARG426-GLY431

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGCTGGAGCA
 GTCTTCGTTTCGCCCCAGGCCGTTGAGAAGCTGTGGGTGACCGTGTACTACGGCGTGCCGTG
 TGGAAAGGAGGCCACCCACCGCTTCTGCGCACGCCAACGGAGCTAC
 GCACAAACGCTGTGGGCCACCCACGCCCTGCGTGGCCACCGACCCCCAACCCCCAGGAGATCGTGT
 GG/GAACCTGAGCGAGAACATTCAACATGTGGAAGAACACATGGTGGAGCAGATGCAACAG
 GACATCATCGCCTGTGGGACAGAGCCTGAAGGCGCTGCGTGAAGCTGGGCAACAGCGTGT
 CACCCAGGCCCTGCCCAAAAGGTGAGCTTCACTTCCACACTGTGGGCCCGGCCCG
 CTCGCCCATCTCTGAAGTGCACGAAAGAATTCACCGGAGCGGCCCTGACCAACAGTGA
 GCACCGTGCAGTGCACCCACGGATCGGCCCTGGTGGAGGACCCAGCTGCTGAGACCGC
 AGCTGCGAGGAGGGAGGGCTGGTGTACCGAGCGAGAACCTACCGACAAACGCCAACGACCAT
 CATCGTCAGCTGAAGGAGAGCGCTGGAGATCAACTCGCAGCCGGCAACACACCCGCA
 AGAGCATCACCATCGGCCCGGCCGCGCTCTACCGCCACGGGAGACATCGGCCGACATCC
 GCCAGGGCCACTGCAACATCAGGGCGAGAGTGGAAACACACCTGAAAGCAGATCGTGTGACC
 AAGCTGAGGCCAGTTGGCAACAAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCC
 CGAGATGTGAGCACGCTTCACTGGCGGAGGTTCTCTACTGCAACAGCACCCAGCT
 GTTCAACAGCACCTGGAACACACCATCGGCCCAACACCAACGGCACCATCACCTGC
 CCTGCCGCATCAAGCATGCAACCCGCGGCCGAGGGCATGTACGCCCGCCCG
 GCGGCCAGATCGCTGCAAGCAACATCAGGCCCTGCTGTGACCCGCGACAGGGCGCAAG
 GAGATCGCAACACACCGAGATCTCGCCCGGCCGAGGAGCATGGCGCAACACTGGCG
 CAGCGCGCTGTAAGTCAAGGGTGGAGAAGTGGAGCTGAGCCCTGGGGCTGGCCCGACCAAG
 CCAAGCGCCGCGTGTGAGCGAGAAGGCCGCGTGAACCTGGGGCTGAGCCCGACCAAG
 TTCTGGGCCGCCGCCACATGGGGCGGCCGAGCCTGACCGTGAACGGTGAAGGGCGC
 CAGCTGCTGAGGGCATCGTCAAGCAGAACACCTGCTGCGGCCATCGAGGCCAGCA
 GCACCTGCTGCACTGTTGAGGCAACCGCAGCTGGGGCATCTGGGGCTGCGAGGGCAAGCTGATCTGC
 ACCACCGCCGTGGCTGGAGACGCCAGCTGGAGCAACAAAGAGCCTGGGACAGATCTGAAACAA
 CATGACCTGGATGGAGTGGAGCGAGATGCAACACTACACCAACCTGATCTACACCCCTGA
 TCGAGGGAGAGGCCAGAACCCAGCAGGGAGAAGAACAGCAGGGAGCTGCTGGAGCTGGAGAACATG
 GGCCAGGGCTGGAACTGTTGCACTCAGCAAGTGGCTGTTGATCATCAAGATCTCAT
 GATCGTGGCGCCCTGGGGCTGCGCATCTGTTACCGTGTAGCATGTAACCGCGT
 GCGCCAGGGCTACAGCCCCCTGAGCTTCCAGACCCGCTTCCCCCCCCCGGGCCCGACCG
 CCCGAGGGCATGAGGGAGGAGGGCGGGAGCGCGACCCGAGCGAGCCCCCTGGTGC
 ACGGCCCTGCTGGCCCTGATCTGGAGCAGCTGGCAGCTGTGCTTACGCTACACCGCC
 TGCGCGACCTGATCTGTGATCTGGCCCGCATCTGGAGCTGCTGTGGGCCGCGCTGGAGG
 CCTGAAAGTACTGGGCAACCTGCTGCACTGATCTGGATCTGGAGCTGAAGAACAGCGCCGT
 AGCTGTTGAGCGCATGCCATGCCATGCCGTGGCGAGGGCACCGACCGCATCATCGAGGTGCC
 CAGGGCATCGGCCGCCCTCTCTGACATCCCCCGGCCATCGGCCAGGGCTTGAGCGCGC
 CTGCTGTAACCTGAG

SEQ ID NO:19 LEU122-SER199; ARG426-LYS432

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTTCGTTGCCACGGCGTGGAGAAGCTGGGGTACCGTGTACTACGGCTGCCGTG
 TGAAGGAGGCCACCAACCTGTTCTGGCCAGCGACGCCAACGGCTACGACACGGAGGT
 GCACAACTGTTGGGCCACCCACGCCCTGCGTGGCCACCGAACCCCAACCCCAAGGAGATCGTCTG
 GGAGAACGTCGACCGAGAACCTCAACATGTGGAAGAACACATGTGAGAGATGACAGAG
 GACATCATCAGCTGTTGGGACCCAGAGCTGAAGGCCCTGCTGAAGCTGGCAACAGCTGAT
 CACCCAGGGCTGCCAACAGGTGAGCTTCGGCCATGCCATCCATCCAACATGCGCCCCCG
 CTTCGCCATCTCGTAAGTGAACGAAAGTGAACCGAAGCCCTGAGCAGGCCCTGACCAACAGTGA
 GCACCGTGCAGTGAACCCAGGCCATCGCCCCCTGGTGTGAGGACCCAGCTGCTGTAACGGC
 AGGCTGGCCAGGAGGGCTGGTGTGATCGCAGCGAACACTTCACCGAACGCCAACGACCAT
 CATCGTGCAGCTGAAGGAGAGCTGGAGATCAACTGCAACCCGGCCCAACAAACACCCGCA
 AGAGCATCACCATTGGCCGGGGGGCTTCACGCGACCGGGAGACATCATGGGAGACATCC
 GCCAGGGCCACTGCAACATCAGGGCAGAGTGGAAACACACCTGAAAGCAGATCGTGTGACC
 AAAGCTGAGGGCCAGTCTCGCGAACAGGACATCTGTTCTGAAGCAGAGCAGCGGGCG
 CGAGTATCGTGTGACAGCTCAACTGCGGGGGAGGTCTTCAACTGCAACAGCAGGAGCT
 GTTCAACAGCACCTGAAACACCATCGCCCCAACAACACCGAACATCACCCCTGC
 CTCGGCGCATACAGAGATCATCACCGCGCGGAGACAGGCAATGTACGCCCGGGCATTC
 GCGGGCGAGATCGGCTGAGCGAACATCACCGGCGCTGCTGACCCCGCAGGGCGCAAG
 GAGATCGACCAACACCAGGAGATCTTCCGGCCGGGGGGGAGCATGCGCAGACAACTGGG
 CAGGGAGCTGTAAGTGAAGGGTGTGAGATCGAGCAGGCTGGGGCTGGCCCAACAGG
 CCAAGCGCCGGTGTGAGCGGGAGAAGCGCGGCGTGGCTGGGGCCATGTCTGGG
 TTCTGGCGCCGGCGGCGAGCACATGGGGCCCGAGCGCTGACCTGACCGTGGAGGGCC
 CAGCTGCTGAGCGGATCGCAGCGAACACCTGCTGCGCATGAGGGCCAGCA
 GCACCTGCTGAGCTGAGCTGGTGTGGGGCATCAAGCAGCTGCAAGGCCGGTGG
 AGGGCTACCTGAAGGAGCAGCAGCTGCTGGGCACTGCGGCTGCAAGGCAACTGTC
 ACCACCGCCGTGCCCTGGAACGCCAGCTGGAGCAACAGAGCTGAGCAATCTGGAACAA
 CATGACCTGGATGGAGTGGGAGGGAGATGCAACACTACACCAACTGATCTACCCCTGA
 TCGAGGAGAGCCAACAGCAGGAGAAGAACAGCAGAGGAGCTGTGAGTGGAGCAAGTG
 GGCCAGCTGTTGGAACCTGGTGCACATCGAACAGTGGCTGTGTAACATCAAAGTCTTCATCAT
 GATCTGTTGGCGCCCTGGTGGGGCTGGCATCGTGTGACCTGCTGAGCATCTGAAACGGCGT
 GCGCCAGGGCTACAGCCCCCTGAGCTTCAAGCAGCCGCTTCCCGCCCCCGCGGGCCCGACCG
 CCCCCGAGGGCATCGAGGAGGGAGGGCGGAGGCCGACCCGGAGCCGAGCGAACGCCG
 ACAGGGCCCTGCTGCCCTGATCTGGGAGCACCTGGCAGCGCTGTGCTGTCAGCTACACCGCC
 TGCGCAGCTGATCTGATCGCCGCCCATGTGAGGAGCTGCTGGGGCGCCGGCTGGAGG
 CCTGGAAGTACTGGGCAACCTGCTGAGTCTGAGGAGCTGAAGAACAGCGCG
 AGCTGTTGAGCGCCATGCCATGCCGTGGGGAGGGACCGAACCGCATCATGAGGGTGGG
 CAGGGCATCGGGCGGCCCTTCCTGCACATCCCCCGCCGATCCGCCAGGGCTTCAGGGCG
 CTGCTGTAACTCGAG

FIG. 22

SEQ ID NO: 20: LEU122-SER199; TRP427-GLY431

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGTGAGCA
GTCTTCGCTTCCGGCAGGGCTGGAGAAGCTGGGGTGACCTGTACTACGGCGTGCCTG
TGAAGAGGAGGCCACCAACCCCTGTTCTGGGCCAGCAGCAGCCAAAGCCTACGACACCGAGG
GCACAACTGTTGGCCACCCACCGCTGCGTGCACCGACCCCAACCCCAAGGAGATCGTGC
GGAGAACGTGACCGAGAACCTCAACATGTGAAAGAACAAACATGTGGAGCAGATGACAG
GACATCATCAGCCTGTGGGAACAGAGCTGGCTGAAGGCTCTGGTGAAGCTGGGCAACAGCTGT
CACCCAGGCCCTGCCCAAGGTGAGCTTCAGGCCATCCCCATCCAACACTGGCCCCGCCGG
CTTCGGCATCTGTAAGTCAACAGAACAAAGTTCACGGCAGGGCCCTGACCAACGTGA
GCACCGTGCAGTGCACCCACGGCTCCGGGGCTGTGGAGCACCAGCTGCTGTCAGCG
AGCCTGGCCGAGGGGGTGGTGTATCGCGAGAACATTCACCGACAAGCGAACAGACCAT
CATCGTGCAGCTGAAGGAGAGCTGGAGATCATGCACCCGCCAACAAACAAACACCGCA
AGAGCATACCATCGCCGGGGCGCCCTACGCCACGGGAGACATCATGGCGACATCC
GCCAGGCCACTGCAACATCAGGGCAGAACAGGACATCTTCAAGCAGAGCAGCGGCCGACCC
AAGCTGAGGGCAGTGGCAACAAAGGACATCTTCAAGCAGAGCAGCGGCCGACCC
CGAGATCTGTATGACAGCTCAACTGGGGGGAGTTCTACTGCAACAGCACCCAGCT
GTTCAAACGAACTGGAAACAAACCATGGCCCAAACAAACCCAAACGGGACCATCACCTGC
CTCGCCGATCAAGCAGATCATCAACCGCTGGGGCGCAAGGGCATGTACGCCCGGCGATCC
GCCGGCAGATCGCTGAGCAGAACATCACGGGCTGCTGTCAGCCCGGAGCGGCCGAG
GAGATCGCAACACCCAGGAGATCTTCCGGCCCGGGCGACATGGCGACAACTGGG
CAGGGAGCTGTAAGTCAAGTGAAGGGTGTGAAGATGAGCCCCCTGGCGTGGCCCCAACAGG
CCAAGCGCCGCTGGTGGAGCAGGGAGAACGGCCTGACCTGGGGCCATGTTCTGGG
TTCTGGGGCCGGGGAGACCATGGGGCCCGACGGCTGACCTGAGCTGCAAGGGCCG
CAGCTGTGAGGGCATGCAAGCAGAACACCTGCTGCGCCCATGAGGGCCAGCA
GCACCTGCTGAGGGCATGACATCGAGCTGGGCTGAGCTGGGCTACAAAGCAGCTGAGGGGG
AGGCTACCTGAAGGACCAAGCAGCTGCTGGGCTATGGGCTGAGCAGGGCAAGCTGATCTGC
ACCAACCGCCGTGCCCTGGAAACGCCAGCTGGAGCAACAAAGAGCTGGACCATGATCTGAAACAA
CATGACCTGGATGGGGAGGGAGGGAGGGAGATCGAACACTACACCAACCTGATCTACACCC
TCGAGGAGAGGCCAACAGCAGGGAGAACAGCAGGGAGCTGCTGAGCTGGCAAGCTG
GGCCAGGCTGTGGAACTGGTTGACATCGAACAGTGGCTGTGGTACATCAAGATCTCAT
GATCTGTGGGGCCCTGGTGGGCTGCGCATGTGTCACCGTGTGAGCATGTCAGCG
GCCAGGGCTACAGCCCGTGAAGCTTCAAGCCGCTTCCCGCCCCCGGGCCCGACCG
CCCCGAGGGCATCGAGGGAGGGGGGGAGGGGGAGGGGGAGGGGGAGGGGGAGGG
ACGGGCTGCTGCCCTGATCTGGAGCACCTGCCAGCCGTGCTGCTGCTGCTG
TGCCTGACCTGATCTGTGAGCTGGGGCCCGCATCTGGAGCTGCTGGGGCCGGGG
CCTGAGAGTACTGGGCAACCTGCTGAGTACTGGGATCCAGGAGCTGAAGAACAGGCC
AGGCTGTTGAGCCATGCCATGCCGTGGCGAGGGCACCGACCGCATCATGAGGGTGG
CAGGGCATCGGGGGCCCTTCTGACATCCCGGCCATCGGCCAGGGCC

FIG. 23

SEQ ID NO:21 LYS121-VAL200; ASN425-LYS432

GAATTGCCACCATGGATGCAATGAAAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTCGTTGCCGACGCCGTGGAGAAGCTGGGGTACCGCTGACTACGGCTGCCGTG
 TGGAGGAGGCCACCAACCCCTGTTCTGGGCCAGCAGCAGCCAAAGCCCTACGACACCGAGGT
 GCACAACTGTTGGGCCACCCACGCCCTGCGTGGCAACGACCCCCAAGGGAGATCGTGC
 GGAGAACGTAACCGAGAACCTCAACATGTGGAAGAACACATGGAGCAGATGACAGAG
 GACATCATCGACCTGTGGGAGCCAAGGCCCTGAGCTGTAAGGGCCCGTGTGATCACCA
 GGCTGCCCCAAGGTGAGCTTGTGAGCCCATCCCCATCTCAACTGTGCCCCCGCCGGCTTC
 CATCTGAATGCAACGAAAGATTCACGGCAAGGCCCTGACCAACGTGAGCAACGG
 TGCACTGACCCACGGCATCCGGCGCTGTGGTGAACCCAGCTGTGTAACGGCACGCTGG
 CGCAGGGAGGGCGTGGTGTATCGCAGCGAGAACCTCAACGCCAACGCCAGACCATCGT
 CAGCTGAAGGAGACGGTGGAGATCACTGACCCCGCCCAACAAACACCCGCAAGAGCAT
 CACCATGGCCCCCGCCGCCCTTACGCCACCGGGGACATCATCGGCAATCGGCCAGGC
 CCACCTGAACTACAGCAGGGCAAAAGACCATGTCAGCAGAGCAGCGGGCGACCCCGAGATC
 AGGCCAGTTCGGCAACAAGACCATGTCAGCAGAGCAGCGGGCGACCCCGAGATC
 GTGATGCAACAGCTCACTGGCGCGCAGTGTCTACTGCAACAGACCCAGTGTCAAC
 AGCACCTGGAAACACCATCGGCCACCATCGACCATCCCTGCCCTGCG
 CATCAAAGAGATCATCAACGCCCAAAGGCCATGTACGCCCCCCCATCGGGCAAGATCG
 CTGAGCAGAACATCAACGGCGTGTGCTGACCGCGACGGCGCAAGGAGATCAGAAC
 CCACCGAGATCTCGCCCGGGCGGCCGACACTGGCGAGCGAGCTGTAC
 AAGTACAAGGTGGTGAAGATCGAGCCCTGTGGCGTGGGCCACCAAGGCCAACGCCCG
 GTGTGCAAGGGCAAAAGCGCCGAGACCTGGCCGCAACTGGCTGGGCTTCTGGCC
 CGGCGACGCAACATGGGGCGCCGAACCTGACCTGAGCTGCAAGGCCGCAAGCTGTGAGCG
 GCATCGTCAGCAGCAGAACACCTGTCGGCGCATCGAGGCCAGCAGCACCTGCTGAG
 CTGACCGTGTGGGACATCAAAGCAAGCTGCAAGGCCCGCTGTGGCTGGAGCGCTACCTGAA
 GGACAGCAGCTGCTGGGCATCTGGGCGTCAAGGGCAAGCTGATCTGACCAACCCGCG
 CCTGGAA CGCCAAGCTGGAGCAA CAAGAGCTGGACCAAGATCTGGAAACACATGACCTGATG
 GAGTGGGA CGCGAGATCGACAACTACACCAACCTGATCTACACCTGATCGAGGAGAGGCCA
 GAACCA CGAGGAGAA GAACGAGCAGGAGCTGTGGAGCTGCAAGTGGCCAGCTGG
 AACTGGTTGCACTCAGCAAGTGGCTGTGGTACATCAAGATCTCATGATGCTGGGCGC
 CTGCTGGGCCCTGCGCATGTTACCGTGTGAGCATCGTGAACCGCTGCGCCAGGGCTAC
 AGCCCCCTGAGCTTCCAGACCCGCTTCCCCGCCCGGGGCCACCGCCCCAGGGCATC
 GAGGAGGA GGCGCGCA GCAGCAGGCCGACCGCAGCAGCAGGCCCTGGTGCAAGGCC
 CCTGATCTGGGACACCTGCGCAGCCTGTGCTGTGCTGAGCTACACCGCTGCGCACTGT
 CCTGATCTGGGCCCGCATCTGTGGAGCTGTGGGCCCGGGCTGGAGGGCCCTGAAGATCTG
 GGGCAA CCTGCTGCACTGGATCAAGGAGCTGAAGAACAGCCGTGAGCTGCG
 CCATCGGCATCGCGTGGCGAGGGCACCGACCGCATCATCGAGGTGGCCAGCGCATCGG
 CGGCCCTTCTGCACTCCCCGCCGATCGGCCAGGGCTGAGCGCCCTGCTGTAAC
 GAG

SEQ ID NO:22 VAL120-ILE201; ILE 424-ALA433

GAATTGCCACATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGTGAGCA
GTCTTCGTTGCCAGGCCCTGGAGAACGCTGGGTGACCGTACTACGGCTGCCGTG
TGGAAAGGAGGCCACCAACCAACCTTGTGCGCCAGCGACGCCAACGGCTACGACACCGAGGT
GCACACAGTGTGGCCACCA CGCCTGGTGCACCGA CCCCCAACCCCCAGGAGATGATGTC
GGAGAACGTCGACCGAGAACCTCAACATGTTGAAAGAACACATGGTGGAGCAGATGACGAG
GACATCATCGCCTGTTGGGACCAAGGCCATGGCTGGGCCCATACCCAGGCTG
CCCCAAGGTGAGCTTGTGAGCCATCCCCATCCTACTATGCGCCCCCGCGCTTGC
GAAGTGCACGACAAGAAAGTCAACGGCAGGCCCTGACCAACGTGAGCACCCTGAGT
GCCACCCACGGCATCCGGCGCTGGTGTGACGCCAGCTGTGTAACGGCAAGCTGGCGAG
GAGGGCGCTGGTGTCCGGAGAACACTTACCGAACAGCCAAAGCCAAACACCCGCAAGAGCATACCA
GAAGGGAGAGCTGGAGAACATGCAACCCGCCAACAAACACCCGCAAGAGCATACCA
TCGGCCCGCCGGCGCCCTTAACGGCACCGCAGACATCGGGCAATCGGGCAATCGGGCCAGGGCCACT
GCAACATCAGCGCCAGAACAGTGGAAACAACACCTGAAAGCAGATGTGACCAAGCTGCAGGC
CAGTTCGCAACAAGGACCATGTTGACAGCAGCAGGGCGGACCCGGAGATCGTAT
GCACAGCTTCAACTGGCGCGGAGCTTCTACTGCAACAGCACCAAGCTGTTCAACAGCAC
CTGAAACAAACCATGGCCCAACAAACCAACCGCAGCACCATCACCCCTGCCCTGCCGATCA
AGCAGATCATCGGGCGGGCCATGTACGGCCCCCCCACATCGGCAGCGCAGATCGCTGAGCAGC
AACATCACGGCGCTGTCGTCGACCCCGAGCGCGCAAGGGAGATCAGCAACACCCAGGAGAT
CTTCGGCGCCGGCGGGCGCAATCGCGCAACTGGCGCAGGGAGCTGTACAAGTACAAGG
TGGTGAAGATTCGACGCCCTGTGGCGCTGGGCCAACAGGGCAAGCGCGCTGGTGAAGCG
GAGAACGCGCCGTGACCTGGCGGACTGTGTCGCTGGCTTCTGGCGCCGGCGCAGCACC
ATGGGGCGCCGCACTGGTGTGGCGCTGGCGGACTGTGTCGAGCGCGCACTGGTGAAGCG
GCAGCAGAACAAACTGTGCGCCGCACTGAGGGCCAGCAGCACCTGTGAGCTGACCGTGT
GGGGCATCAAGGAGCTGCAAGCCCGCGCTGGCTGGCGCTGGAGCGCTACCTGAAGGACCAAG
CTGCTGGGCATCTGGGCTGCAAGGGCAACTGTGATCTGACCAACCCCGTGGCTGAACGCC
AGCTGGAGAACAAAGGCCCTGAGCACAGAACATGACCTGGAGCTGGAGCTGGAGCG
CGAGATCGACAACATACCAACACTGATCTGACCCCTGATCGAGGAGAGGCCAGAACCGAGG
AGAACAGGAGCAGAGCTGCTGGAGCTGAGACAAGTGGGCCAGCTGTGGAATGGTTCGAC
ATCAGCAAGTGGCTGTGGTACATCAAGATCTTACATGATGCTGGCGGCCCTGGCTGG
CGCATCGTGTGACCGTGTGAGCATGTGAAACCGCGTGCAGCCAGGGCTACAGCCCTGAGC
TTCCAGACCCGCTTCCCGCCCCCGCGGGCCCGACCGCCCGAGGGCATCGAGGAGGG
CGGGAGGCCGACCGGAGCGCAGAGGCCCTGGTGCAGGGCTGCTGGCCCTGATCTGG
ACGACTGCGCACGCTGTGCTGTGAGCTGACCCGCGTGCAGGACCTGATCTGAGCCG
CCCGCATCTGTGGAGCTGCTGGGCCGGCGCGCTGGAGGCCCTGAGATCTGGGCCAACCTG
CTGAGTACTGGATCAGGAGCTGAAAGAACAGCGCGTGAAGCTGCTGACGCCATGCCAT
GCCGTTGGCGGAGGGCACCGACCGCATCGAGGTGGGCCAGCGCATGGCGCCCTGCTGAACCG
GCACATCCCCCGCGCATCCGCCAGGGCTTCGAGCGCGCCCTGCTGAACCTGAG

FIG. 25

SEQ ID NO:23: VAL120-ILE201B; ILE424-ALA433

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGTGGAGCA
 GTCTTCGTTGCCACGGCGTGGAGAACGCTGGGGTACGGCTGACTACGGCTGCCGTG
 TGGAAAGGAGGCCACCAACCCACCTCTTCTGGGCCAGCGACGCCAACGGCTACGACACCGAGGT
 GCACAACTGTGGGGCCACCCACCGCTGCCGCCCCACGGAGATGACAG
 GGAGAACGTGACCGAGAACCTCAACATGTGGAAGAACACATGGTGGAGCAGATGACAG
 GACATCATCGCCTGTTGGGACCAAGGCTGAAGGCTGCTGTCGGGGCATACCCAGGGCTGC
 CCCAAGGTGAGCTTGAGGCCATCCCCATCACTACTGCCCCCGCCGGCTTGCGCATCTG
 AAGTGCACAGAACAGATTCAACGGCAGCGGCCCCCTGCACCAACGTGACGACCGTGCAGTG
 CACCCACGGCATCGGCCCGCTGGTGAACGACCCAGCTGCTGTAACGGCAGGCTGGCAGG
 AGGGCGTGTGATCCGACGGAGAACCTACCGACAACGCCAACGACCATCATGTCAGCTG
 AAGGAGACGGTGGAGAACATCACTGACCCGCCAACAACACCCGCAAGAGCATCCACAT
 CGGCCCCCGGCCGCTTCTACGGCACCGGACATCATCGCGACATCGCGCAGGCCACTG
 CAAACATCAGCGCGAACAGTGGAAACACACCTGTGAAGCAGATGTCGACCAAGCTGCAGGCC
 AGTTCGGCAACAGACCATCGTGTCAAGCAGAGCAGCGAGCGGGGGGAGATCTGTGATG
 CACAGCTCACTGGCGCGGAGTCTCTACTGCAACGACCCAGCTGTCACAGCACC
 TGGAAACACACCATGGGCCAACACACCAACCGCACCCATACCTGCGCCGATCAA
 GCAGATCATCGCGCGGCCATGTCAGGCCCTCCATCGCGGCCAGATCGCTGCAAGCAGA
 ACATCACCGGCCCTGCTGCTGACCGCGACGGCGCAAGGAGATCAGCAACACCCAGGAGATC
 TTCCGGCCCGGGCGGCCGACATGCGCAGAACACTGGCAGGGAGCTGTACAAAGTACAAGG
 GTGTAAGATCGAGCCCTGCGCTGGGCCACAGGGCAAGCGCCGCGTGGTGCAGCG
 AGAAAGCGGCCGCTGACCTGGGCCATGTTCCTGGGCTTCTGGCCGGCGGCCAGACCA
 TGGCGCCCGCAAGCTGACCTGAGCGCCATCGAGGCCAGCAGCTGTCAGCGGATCG
 CAGCAGAACACACTGCGGCCATCGAGGCCAGCAGCACCTGTCAGCTGACCGCTGTG
 GGGCATCAAGCAGCTGCAAGGCCCGCTGCTGGCCGTGGAGCCTACCTGAAGGACAGCAG
 TCGTGGCATCTGGGCTGCAAGGGCAAGCTGATCTGACACCCCTGATCGAGGAGAGCCAGAACGAGGA
 GAGATCGACAACATACCAACACTGATCTGACACCCCTGATCGAGGAGAGCCAGAACGAGGA
 GAAGAACGAGCAGGAGCTGTCAGCTGGAGCTGACAAAGTGGCCAGCCTGTTGAACTGTTGAC
 TCAGCAAGTGGCTGTGGTACATCAAGATCTTCATCATGATCTGTCGGGGCCCTGGTGGGCTGC
 GCATCTGTTCACCGCTGCTGACGATCTGTAACCGCGTGGCCAGGGCTACGCCCTGAGCT
 TCCAGACCCGCTTCCCGCCCCCGGCCGGGCCCCGACGGCCCGAGGGCATCGAGGAGGG
 GGCAGGCGGCCGGAGCGCGACCGCAAGCGCCCCCTGTTGCAAGGGCTCTGGCCCTGATCTGG
 CGACCTGCGCACCTGTGCTGTAGCTACCCCGCTGCGGCCAGCTGATCTGATGCC
 CGCGCATCTGTCAGGCTGCTGGGCCGGCGCGCTGGAGGGCCCTGAGATACTGGGCAACCTGC
 TCGAGTACTGGATCAGGAGCTGAAGAACACGCCGCTGAGCTGTTGACGCCATGCCATC
 GCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCCAGCGCATCGGCCGCCCTG
 GCACATCCCCGGCGCATCGGCCAGGGCTTCGAGCGGCCCTGTCAGCGAG

FIG. 26

SEQ ID NO:24 VAL120-THR202; ILE424-ALA433

GAATTCGCCACCATGGATGCAATGAAGAGGGCTGCTGTGCTGCTGTGCTGTGGAGCA
 GTCTTCGTTGCCAGCGCGTGGAGAAGCTGGGGTGAAGCTGTTACTACGGCGTGCCTG
 TGGAAAGGAGGCACCAACCCACCTGTTCTGCGCCAGCGAACGCCAAGGCCTACGACACCGAGGT
 GCACAACGCTGTTGGGCCACCCACGCCCTGCGTGCCTGCCAGGACCCCAACCCCCAGGAGATCTGCTG
 GGAGAACGTCGACCGAGAACCTCAACATGTGGAAAGAACACATGTGGAGCAGATGACAGAG
 GACATCATCGACCTGTGGGACCAAGAGCTGGCTGGCGCGGCCACCCAGGCTG
 CCCCAGGGTGAACCTTCGAGCCCATGCCCTACCAACTGTGGGCCCGGCCGCTTCGCCATCCT
 GAAGTGCACAGACAAGAAGTCAAGCGAGCGGGCCCTGCACCAACGTGAGCACCGTGCAGT
 GCACCCACGGCATTCGGCCCTGTTGAGCACCCAGCTGCTGAAACGGCAGGCTGGCCAG
 GAGGGCGCTGGTATCGCGACCGAGAACCTTCACCGGAAACGCCAACGACCATATCGTCAGCT
 GAAAGGAGACGGTGGAGATCAACTGCAACCCGGCCAAACAAACACCCGCAAGAGCATACCA
 TCGGCCCCGGCGGCCCTACCGGCCACGGCAATCATGGCGACATCGCCAGGCCACT
 GCAAACATCAGCGGCGAGAAAGTGGAAACACACCTGAGAGCAGATGTCGACCAAGCTGCAAGGCC
 CAGTGTGCAACAAAGGACATCTGTTCAAGCAGCAGCAGGCTGTAACAGTACAAGG
 GCACAGCTTCAACTGGCGGCCGAGTTCTACTGCAACAGCACCAAGCTGTTCAACAGCAC
 CTGGAAACAAACCATGGGCCAACAACACCAACGGCACCATCACCTGCCCTGCCCATCA
 AGCAGATCATCGGGCGGCCATGTCACCCCCCCCACATCGGGCGCAGATCGCTGAGCG
 AACATCACGGGCTGCTGTCGACCCCGCAGGGCGCAAGGAGATCAGCAACACCCAGAGAT
 CTTCGTTCAAGTGTGAGCCCCCGGGCGCATGGCGACACAAGTGGCAGCGAGCTGTAACAGTACAAGG
 TGGTGAAGATCGAGCCCCCGGGCGCATGGCGACACAAGTGGCAGCGAGCTGTAACAGTACAAGG
 GAGAAAGCAGCGCGTGTGGCGGCGATGTCTCTGGGCTTCCTGGCGCCGCCAGCAGC
 ATGGCGCCCGAGCGTGTGGCGCATGGCTGACCGTGCAGGCCAGCTGCTGAGCGGCCATGTGCA
 GCAGCAGAACACACTGCTGCCGCATCGAGGCCAGCAGCACCTGCTGAGCTGACCGTGT
 GGGGCATCAAGCAGCTGAGGGCGCCGCTGTCAGGGCTGAGCGCTACCTGAAAGGACAGCAG
 CTGCTGGGCATCTGGGCTGCAAGCTGATCTGCAACCCGCCGTGCCCTGGAACGCC
 AGCTGGAGACAACAAAGAGCTGGACCAGATCTGAAACAAACATGACCTGGATGGAGTGGAGCG
 CGAGATCGACAACATACACCAACCTGATCTACACCTGATCGAGGGAGAGCCAGAACACAGCAGG
 AGAAAGAACAGGAGCGAGCTGTGGAGCTGGCAAGTGGGCCAGCTGTGAAACTGTTTGAC
 ATCAGCAAATGCGCTGTGTCATCTAACAGATCTTCATCATGATCTGTTGGGGCTGTGGGGCTG
 CGCATCTGTTACCGTGTGAGCATCTGTAACCCGGCTGCGCCAGGGCTACAGCCCCCTGAGC
 TTCCAGACCGCTTCCCAGCCCGGGCGACCGCAGCAACCCCTGTGCAAGGGCTCTGGCCCTGAGCTGG
 CGGGAGCGGCCACCCGGAGCCGAGCGACGCCAGGGCTACAGGGCTCTGGCCCTGAGCTGG
 ACGACCTGCGCAGCTGTGCTCTGAGCTGACCTGATCTGATCTGACCTGCG
 CCGCATCTGAGCTGTCAGCTGACCTGAGCTGAAAGAACAGCAGCTGAGCTGAACTGTTGCACTG
 CTGCACTGAGCTGAGCTGTCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG
 GCGCTGGCGAGGGCACCACCGCATCTGAGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG
 GCACATCCCCGGCGCATCGGCCAGGGCTTCGAGCGGCCATGCGACCGCCATGCCATC

FIG. 27

SEQ ID NO:25 VAL127-ASN195

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGTGAGCA
 GTCTTCGTTGCCAGGCCAGGGAAAGCTGTGGGTGACCGTGTACTACGGCTGCCGTG
 TGGAAAGGAGGCCACCCACCCACCTGTTCTGGGCCAGCGACGCCAAGGCCCTACGACACCGAGGT
 GCACAAACGTGTGGGCCACCCACGCCCTGCGTGCACCCACGGACCCAAACCCCAAGGAGCGAGATGCG
 GGAGAACTGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGGTGGAGCGAGATGCG
 GACATCATCGCCTGTGGGACCCAGAGCTGAAGGCTGTGAGCTGACCCCTGTGCGTG
 GGGGCAGGGAACTGCAACACCCAGCTGATCACCCAGGCCCTGCCCAAGGTGAGCTTGCAGCC
 CATCCATCCATCCACTGTGCCCGCCCGCCCTGCCCATCTGAAGTGTGCAACGACAAGAAGIT
 CAACCGCAGCGGCCCTGACCAACGTGAGCACCGTGTGAGTGCACCCACGGCATCGCCCG
 TGGTGAGCACCAAGCTGCTGAAACGGCAGCTGCGAGGGAGGGCTGTTGATCCGAGC
 GAGAACTTCAGCAGCAACGCCAACAGGACATCATGGCAGCTGAAGGAGAGCTGAGGATCAA
 CTGCAACCGCCCAACAAACACCCGCAAGGACATCACCATCGGCCCGGCCCTCTA
 CGCCACCGCGACATCATGGCAACATGCCGCAAGGCCCATCTGAACATCAGCGGCGAGAAGT
 GGAAACACCTCGAAAGCAGATCGTGAACAGCTGAGGCCAGTGGCAACAAAGACCATC
 GTGTTCAAGCAGAGCAGCGGGGGACCCAGATCGTGTGATGCACAGCTTCAACTCGCG
 CGAGTCTCTACTGCAACAGCACCCAGCTGTCACAGCACCTGAAACAAACCATCGCC
 CAACAAACCAACCGCACCATCACCTCGCTGCCATCACAGCAGATCATCACCGCTGG
 AGGAGGTGGCAAGGCGATCTAGCCCCCCCCATCGGGGCCAGATCGCTGAGCAGCAAC
 ATCACCGCCCTGCTGTAACCGGGCACAGGGAGATCAGCAACACCCACCGAGATCTT
 CCGCCCCGGCGGGCGCATGGCGACAAGTGGCAGCGACTGTAACAGTACAAAGGTGG
 TGAAGAGTGAAGGCCCCCTGGGCAACCGGCAAGGGCCCGTGTGCAAGCGCAG
 AAGGGCGCGCTGACCTGGGCCCATGTTCTGGGCTTCTGGGCCCGGCCAGCACCATG
 GGCGCCCGCAGCTGACCGTGCAGGCCAGCTGCTGAGCGGATCGTGCAGCA
 GCAGAACACCTGCTGCGGCCATCGAGGCCAGCAGCACCTGCTGAGCTGACCTGTGG
 GCATCAAGCAGCTGAGGCCCGCTGCTGGCGTGTGAGCGTACCTGAAAGGACCGAGCTG
 CTGGGCACTGGGCTGAGCGCAAGCTGATCTGACCAACCCGGCTGGCTGGAAACGCCAG
 CTGGAGAACAAAGAGCTGGACCATGTTGAAACATGACCTGGATGGAGTGGAGGCCAG
 AGATGACAACCTACACCAACCTGATCTACACCCGATCGAGGAGAGGCCAGAACAGCAGGAG
 AAGAACGAGCAGGAGCTGCTGGAGCTGAGAACAGTGGGCCAGCCCTGGAACCTGGTGCACAT
 CAGCAAATGGCTGTGATCATCAAGATCTTACATCATGATCTGTTGGGCCCTGTGGGCCCTGCG
 CATCGTGTACCGTGTGAGCATCGTGAACCCGCTGCGCCAGGGCTACAGCCCCCTGAGCTT
 CCAGACCCGCTTCCCCGCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
 GCGAGCGCACCGCACCGCACAGCCCCCTGGTCAGGCCGAGGGCATCGAGGGAGGGGG
 GACCTGGCAAGCCTGTGCTGTCACTACCAACCCGCTGGCGACCTGATCTGATCTGGGAC
 CGCATCTGGAGACTGCTGGGCCGGCGGGCTGGAGGCCCTGAAGTACTGGGCAACCTGCT
 GCAGTACTGGATCCAGGAGCTGAAGAACAGGCCGTGAGCCCTGTTGACGCCATGCCATCG
 CGTGGCCGAGGGCACCGACCGCATCATGAGGTGGCCAGCGCATCGGCCCGCTTCTGC
 ACATCCCCCGGCCATCCGCCAGGGCTTGAGCGGCCCTGCTGTAACTCGAG

FIG. 28

SEQ ID NO:26 VAL127-ASN195; ARG426-GLY431

GAATTGCCACCATGGATGCAATGAAGAGGGCTCTGCTGTGCTGCTGTGAGCA
 GTCTTCGTTGCCAGGCCAGGGCTGGAGAAGCTGGGTGACGGTGTACTACGGCTGCCGTG
 TGGAAAGGAGGCCACACCCACCCCTGTTCTGGGCCAGCAGCAGCCAAAGGCTACAGCACCGAGGT
 GCACAAACGTTGCGCACCCACCGCTGCGTGCCTACCGACCCCAACCCCAAGGAGATCGTGTCT
 GGAGAACGTTGACCGAGAACCTCAACATGTGAAAGAACACATGTGAGCAGATGACAGAG
 GACATCATCGCCTGTGGGACCAAGGGCTGAAGGCTCTGGTGAAGCTGACCCCCCTGTGGGTG
 GGGCAGGGAACTGCAACACCAAGCTGATCACCCAGGCTGCCAACGGTGAAGCTTGAGCC
 CATCGCCATCCACACTCGGCCCGCCGCTCTGCCCATCTGAATGCAACGACAAGAATT
 CAACGGCAGGGCCCTGCAACACCGTGAAGCACCGTCACTGCAACCCACGGCATCUCGGCC
 TGGTGAGCACCCAGCTGCTGAAACGGCAGCTGCCAGGGCTGAGGAGGGCTGTGATCCGGCAGC
 GAGAACATTCGCGAACACGCCAACGACCATCATCGCCAGGCTGAAGGAGAGGGTGAAGATCAA
 CTGCAACCGGCCAACAAACACCCGCAAGAGCATCACCATCGGCCAGGCCCCTGCAACATCAGCGCGAAGA
 CGCCACCGGCGACATCATCGCGACATCCGCCAGGCCCCTGCAACATCAGCGCGAAGA
 GTGGAACACCGGCTTAACGGCAGCATCGTGAACCAAGGGCTGAAGGAGGGCTGAGGAGACCCAT
 GTGTTCAAGCAGAGCAGCGCCGCCAGGACAGATCGTGAAGCTGCAAGCTTCACTGGCG
 CGAGATTCTCTACTGCAACAGCAGGCCACCGTGTCAACAGCACCTGGAACAACACCATCGGCC
 CAACAAACACCAAGGCCACCATCACCTGCTGCCGATCACAGCAGATCATCACCGGCC
 GCGCGCAAGGCCATGTACGCCCGCCCATCGCCGCCAGATCGCTGAGCAGAACATCACC
 GGCGCTGTGACCGGCCAGGCGGCCAGGAGATCGCAACACCCAGGAGATCTCGGCC
 CGGGGGCGCGACATCGCCGACAATCGCCGAGCGAGCTGTAACAGTACAAGGTTGTAAG
 ATCGAGCCCTGCGGCTGGCCACCAAGGCGAACAGCGCCGCTGGTGCAGCGCGAAGAGCG
 CGCCGTGACCTGGGCCCATGTGCTGGGCTTCCTGGGCCCGCCCGCAGCACCATGGGCC
 CGCGCAGCCTGACCCGTGACCGTGCAGGCCCGCCAGCTGCTGAGCGCATCGTCAGCAGCAGA
 ACAACACTGCTGCCGCACTGAGGCCCGACAGCACCTGCTGCAAGCTGACCTGGTGTGGGCATCG
 AAGCAGCTGCGGCCGCGCTGCTGGCGTGGAGCGCTACCTGAAAGGACAGCAGCTGTTGG
 CATCTGGGGCTGCAAGGGCAAGCTGTACTGCAACACCCGCTGCCCCGAACGCGACTGGA
 GCAACAAAGAGGCTGGACAGATCTGGAACAACATGACCTGGATGGAGTGGGAGGCCGAGATC
 GACAACACTACACCAACCTGATCATCACCTGATCGAGGAGAGGCCAGAACCGAGAGAAGA
 CGAGCAGGAGCTGCTGGAGCTGACAAAGTGGCCAGCCTGCTGAGACTGGTTCGACATCGA
 AGTGGCTGTGTTACATCAAGATCTCATCATGATCTGGGGCCGCTGTTGGGCCATCG
 TGTTCACCGTGTGAGCATCGTAACCGCGTGTGCCAGGGCTACAGCCCCCTGAGCTTCCAGA
 CCGCTTCTCCCGCCCCCGGCCGGCCGGCCAGCAGGCCAGGGCTACAGCCCCCTGAGCTTCCAGA
 CGCGACCGCGACCCGAGCACCCCCCTGCTGACCGCCCTGCTGGCCCTGATCTGGGACGACTG
 CGCAGCCCTGTCGCTGCTGAGCTACACCCCTGCGGACCTGATCTGATGCCGCCGATC
 GTGGAGCTGCTGGGCCGCCGCGCTGGAGGGCTGAAGTACTGGGCAACCTGTCAGTA
 CTGGATCAGGAGCTGAAGAACACGCCATCATCGAGGTGGCCAGCGCATCGGCCGCTCTGCACATCC
 CGCGCCGACCCAGCCGATCGAGGCTCTGAGCAGGCCATCGCCATCGGCCGCTCTGCACATCC

SEQUENCE LISTING

<110> Chiron Corporation

<120> MODIFIED HIV ENV POLYPEPTIDES

<130> 1605.100

<140>

<141>

<160> 26

<170> PatentIn Ver. 2.0

<210> 1

<211> 856

<212> PRT

<213> Human immunodeficiency virus

<400> 1

Met	Arg	Val	Lys	Glu	Lys	Tyr	Gln	His	Leu	Trp	Arg	Trp	Gly	Trp	Arg
1				5					10				15		

Trp	Gly	Thr	Met	Leu	Leu	Gly	Met	Leu	Met	Ile	Cys	Ser	Ala	Thr	Glu
									25				30		

Lys	Leu	Trp	Val	Thr	Val	Tyr	Tyr	Gly	Val	Pro	Val	Trp	Lys	Glu	Ala
									40				45		

Thr	Thr	Thr	Leu	Phe	Cys	Ala	Ser	Asp	Ala	Lys	Ala	Tyr	Asp	Thr	Glu
									55				60		

Val	His	Asn	Val	Trp	Ala	Thr	His	Ala	Cys	Val	Pro	Thr	Asp	Pro	Asn
65									75				80		

Pro	Gln	Glu	Val	Val	Leu	Val	Asn	Val	Thr	Glu	Asn	Phe	Asn	Met	Trp
									85				90		95

Lys	Asn	Asp	Met	Val	Glu	Gln	Met	His	Glu	Asp	Ile	Ile	Ser	Leu	Trp
									100				105		110

Asp	Gln	Ser	Leu	Lys	Pro	Cys	Val	Lys	Leu	Thr	Pro	Leu	Cys	Val	Ser
									115				120		125

Leu	Lys	Cys	Thr	Asp	Leu	Lys	Asn	Asp	Thr	Asn	Thr	Asn	Ser	Ser	Ser
									130				135		140

Gly	Arg	Met	Ile	Met	Glu	Lys	Gly	Glu	Ile	Lys	Asn	Cys	Ser	Phe	Asn
									145				150		160

Ile	Ser	Thr	Ser	Ile	Arg	Gly	Lys	Val	Gln	Lys	Glu	Tyr	Ala	Phe	Phe
									165				170		175

Tyr	Lys	Leu	Asp	Ile	Ile	Pro	Ile	Asp	Asn	Asp	Thr	Thr	Ser	Tyr	Lys
									180				185		190

Leu Thr Ser Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys Val
 195 200 205

Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala
 210 215 220

Ile Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr
 225 230 235 240

Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val Ser
 245 250 255

Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val Val Ile
 260 265 270

Arg Ser Val Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln Leu
 275 280 285

Asn Thr Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg
 290 295 300

Lys Arg Ile Arg Ile Gln Arg Gly Pro Gly Arg Ala Phe Val Thr Ile
 305 310 315 320

Gly Lys Ile Gly Asn Met Arg Gln Ala His Cys Asn Ile Ser Arg Ala
 325 330 335

Lys Trp Asn Asn Thr Leu Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln
 340 345 350

Phe Gly Asn Asn Lys Thr Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp
 355 360 365

Pro Glu Ile Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr
 370 375 380

Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp
 385 390 395 400

Ser Thr Glu Gly Ser Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu
 405 410 415

Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Lys Val Gly Lys
 420 425 430

Ala Met Tyr Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn
 435 440 445

Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Asn Ser Asn Asn Glu
 450 455 460

Ser Glu Ile Phe Arg Pro Gly Gly Asp Met Arg Asp Asn Trp Arg
 465 470 475 480

Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val
 485 490 495

Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg Ala
 500 505 510

Val Gly Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser
 515 520 525
 Thr Met Gly Ala Ala Ser Met Thr Leu Thr Val Gln Ala Arg Gln Leu
 530 535 540
 Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu
 545 550 555 560
 Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu
 565 570 575
 Gln Ala Arg Ile Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu
 580 585 590
 Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val
 595 600 605
 Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Glu Gln Ile Trp Asn
 610 615 620
 His Thr Thr Trp Met Glu Trp Asp Arg Glu Ile Asn Asn Tyr Thr Ser
 625 630 635 640
 Leu Ile His Ser Leu Ile Glu Glu Ser Gln Asn Gln Glu Lys Asn
 645 650 655
 Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp
 660 665 670
 Phe Asn Ile Thr Asn Trp Leu Trp Tyr Ile Lys Leu Phe Ile Met Ile
 675 680 685
 Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile
 690 695 700
 Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr His
 705 710 715 720
 Leu Pro Thr Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu Glu
 725 730 735
 Gly Gly Glu Arg Asp Arg Asp Arg Ser Ile Arg Leu Val Asn Gly Ser
 740 745 750
 Leu Ala Leu Ile Trp Asp Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr
 755 760 765
 His Arg Leu Arg Asp Leu Leu Ile Val Thr Arg Ile Val Glu Leu
 770 775 780
 Leu Gly Arg Arg Gly Trp Glu Ala Leu Lys Tyr Trp Trp Asn Leu Leu
 785 790 795 800
 Gln Tyr Trp Ser Gln Glu Leu Lys Asn Ser Ala Val Ser Leu Leu Asn
 805 810 815
 Ala Thr Ala Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val
 820 825 830

Val Gln Gly Ala Cys Arg Ala Ile Arg His Ile Pro Arg Arg Ile Arg		
835	840	845

Gln Gly Leu Glu Arg Ile Leu Leu	
850	855

<210> 2
<211> 847
<212> PRT
<213> Human immunodeficiency virus

<400> 2			
Met Arg Val Lys Gly Ile Arg Lys Asn Tyr Gln His Leu Trp Arg Gly			
1	5	10	15

Gly Thr Leu Leu Leu Gly Met Leu Met Ile Cys Ser Ala Val Glu Lys		
20	25	30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr		
35	40	45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val		
50	55	60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro			
65	70	75	80

Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys		
85	90	95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp		
100	105	110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu		
115	120	125

His Cys Thr Asn Leu Lys Asn Ala Thr Asn Thr Lys Ser Ser Asn Trp		
130	135	140

Lys Glu Met Asp Arg Gly Glu Ile Lys Asn Cys Ser Phe Lys Val Thr			
145	150	155	160

Thr Ser Ile Arg Asn Lys Met Gln Lys Glu Tyr Ala Leu Phe Tyr Lys		
165	170	175

Leu Asp Val Val Pro Ile Asp Asn Asp Asn Thr Ser Tyr Lys Leu Ile		
180	185	190

Asn Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe		
195	200	205

Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu		
210	215	220

Lys Cys Asn Asp Lys Lys Phe Asn Gly Ser Gly Pro Cys Thr Asn Val			
225	230	235	240

Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln
 245 250 255

Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Gly Val Val Ile Arg Ser
 260 265 270

Glu Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Lys Glu
 275 280 285

Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser
 290 295 300

Ile Thr Ile Gly Pro Gly Arg Ala Phe Tyr Ala Thr Gly Asp Ile Ile
 305 310 315 320

Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser Gly Glu Lys Trp Asn
 325 330 335

Asn Thr Leu Lys Gln Ile Val Thr Lys Leu Gln Ala Gln Phe Gly Asn
 340 345 350

Lys Thr Ile Val Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile Val
 355 360 365

Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser Thr
 370 375 380

Gln Leu Phe Asn Ser Thr Trp Asn Asn Thr Ile Gly Pro Asn Asn Thr
 385 390 395 400

Asn Gly Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Arg
 405 410 415

Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Arg Gly Gln
 420 425 430

Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly
 435 440 445

Gly Lys Glu Ile Ser Asn Thr Thr Glu Ile Phe Arg Pro Gly Gly
 450 455 460

Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val
 465 470 475 480

Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val
 485 490 495

Val Gln Arg Glu Lys Arg Ala Val Thr Leu Gly Ala Met Phe Leu Gly
 500 505 510

Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Arg Ser Leu Thr Leu
 515 520 525

Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Asn
 530 535 540

Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr
 545 550 555 560

Val	Trp	Gly	Ile	Lys	Gln	Leu	Gln	Ala	Arg	Val	Leu	Ala	Val	Glu	Arg
565														570	575
Tyr	Leu	Lys	Asp	Gln	Gln	Leu	Leu	Gly	Ile	Trp	Gly	Cys	Ser	Gly	Lys
	580													585	590
Leu	Ile	Cys	Thr	Thr	Ala	Val	Pro	Trp	Asn	Ala	Ser	Trp	Ser	Asn	Lys
	595													600	605
Ser	Leu	Asp	Gln	Ile	Trp	Asn	Asn	Met	Thr	Trp	Met	Glu	Trp	Glu	Arg
	610													615	620
Glu	Ile	Asp	Asn	Tyr	Thr	Asn	Leu	Ile	Tyr	Thr	Leu	Ile	Glu	Glu	Ser
	625													630	635
Gln	Asn	Gln	Gln	Glu	Lys	Asn	Glu	Gln	Glu	Leu	Leu	Glu	Leu	Asp	Lys
	645													650	655
Trp	Ala	Ser	Leu	Trp	Asn	Trp	Phe	Asp	Ile	Ser	Lys	Trp	Leu	Trp	Tyr
	660													665	670
Ile	Lys	Ile	Phe	Ile	Met	Ile	Val	Gly	Gly	Leu	Val	Gly	Leu	Arg	Ile
	675													680	685
Val	Phe	Thr	Val	Leu	Ser	Ile	Val	Asn	Arg	Val	Arg	Gln	Gly	Tyr	Ser
	690													695	700
Pro	Leu	Ser	Phe	Gln	Thr	Arg	Phe	Pro	Ala	Pro	Arg	Gly	Pro	Asp	Arg
	705													710	715
Pro	Glu	Gly	Ile	Glu	Glu	Glu	Gly	Glu	Arg	Asp	Arg	Asp	Arg	Ser	
														725	730
Ser	Pro	Leu	Val	His	Gly	Leu	Leu	Ala	Leu	Ile	Trp	Asp	Asp	Leu	Arg
														740	745
Ser	Leu	Cys	Leu	Phe	Ser	Tyr	His	Arg	Leu	Arg	Asp	Leu	Ile	Leu	Ile
														755	760
Ala	Ala	Arg	Ile	Val	Glu	Leu	Leu	Gly	Arg	Arg	Gly	Trp	Glu	Ala	Leu
														770	775
Lys	Tyr	Trp	Gly	Asn	Leu	Leu	Gln	Tyr	Trp	Ile	Gln	Glu	Leu	Lys	Asn
														785	790
Ser	Ala	Val	Ser	Leu	Phe	Asp	Ala	Ile	Ala	Ile	Ala	Val	Ala	Glu	Gly
														805	810
Thr	Asp	Arg	Ile	Ile	Glu	Val	Ala	Gln	Arg	Ile	Gly	Arg	Ala	Phe	Leu
														820	825
His	Ile	Pro	Arg	Arg	Ile	Arg	Gln	Gly	Phe	Glu	Arg	Ala	Leu	Leu	
														835	840
<210> 3															
<211> 2310															
<212> DNA															
<213> Artificial Sequence															

<220>

<223> Description of Artificial Sequence: Val120-Ala204

<400> 3

<210> 4

211 2316

212 DNA

<213> Artificial Sequence

220

<223> Description of Artificial Sequence: Val120-Ile201

<400> 4

gaattgcgc ccatggatgc aatggagaga gggtctgtgt gtgtgtgtgt gctgtgtgg 60
gcgatcttcg ttccgtcccg ccggctggaa aagctgtgg tgaccgtgtt cracggctgg 120
ccctgttggc aaggccccac caccacccctt tttctggccca gggcggccaa ggccatccac 180
acccggggrc acaacgtgtc ggccacccac gctgtgtgtc ccacccggaa caaccggcc 240
gagatgtttc tgagaaacgtt gagccggatcc ttcaatgtt ggaagaaacaaat ctatgttgtgg 300
catatgttccat aqacatcatc caccctgtgtt qaccacatcc taacccctt ctgtggccac 360

<210> 5
<211> 2322

<212> DNA

<213> Art

6223

2020 Description of National Diabetes Public Health Activities

4009-3

aacaccaaca gcacccatca cctggccctgc cgccatcaagc agatccatcaa ccggctggcag 1020
 gaggtggggca aggccatgt cgccccccct atccggcgcc agatcccgct cagcagcaac 1080
 atcacccccc ttcgctgtcg ccggcgcggc ggcaaggaga tcagcaaac caccggatcc 1140
 ttccggcccg ggggggcgc catgcgcgc aactggcgca gcaaggctgt caagtaaca 1200
 gtgtgtgaaga tcgagccct ggggctggcc cccaccaagg ccaaqgcgcg cgtgtgtcg 1260
 cggcggaga gcccgcgt cctggggcgcc atgttccctgg gtttccctggg ccggccggc 1320
 agcaccatgg ggccggcc cctgacccctg accgtgcagg cccgccagg gctgtggc 1380
 atcggtcgag aggcaaca cctgtcgcc gcccgcgg cccaaggcaga cctgtgtcg 1440
 ctgaccgtgt ggggcatcaa gcagctgcag gcccgcgtgc tggccgtgg gcgttacctg 1500
 aaggaccacg agctgtcg catctgggc tgcaggcgca agctgtatcg caccaccgc 1560
 gtggccctgt ggggccggc gacggcgcgt gagaacaaca ggccgtggacc agatctggaa caacatcgac 1620
 tggatggatg ggggggcgc gatgcacca tacaccaacc tgcgttacac cctgtatcg 1680
 gagagccaga accaggdga gaagaacgcg caggagatgc tggagctgg caagtggggc 1740
 agcctgtgg actgttgc catcgaac tggctgtgg atcataa atcttcatcat 1800
 atlcgtgggg gcctgtgg cttcgcgc gttgcaccc tgcgttaccc tgcgttgc 1860
 gtgcgcggc ggggccggc cttcgcac cttcgcgtt cagacccctg tcccccggcc ccggggccccc 1920
 gaccggccccc aggggatcg ggggggggc ggccggggc ccggccggcc cagcagcccc 1980
 ctgtgtcgac gcctgtgg cttcgcgtt ggccggatgc gcggccctgt ctgtgtcag 2040
 taccacccgc tgcgtgcact gttcgcgtt ggccggccgc tgcgtggat gttggccgc 2100
 cgcggctggg aggccctgaa gtactgggg aaacctgtgc agactctggat ccaggactg 2160
 aagaacacgc cgccgtggat gttcgcgtt atccgcac ccgtggccgc ggccacccgc 2220
 cgcacatcg agggtggccca ggccatcg ccggccgtcc tgcacatccc ccggccgcac 2280
 cgcaggcgt tcgagcgcc cctgtgtta ctcgagatgc ct 2322

<210> 6

<211> 2328

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Lys121-Val200

<400> 6

gaatttcggca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt gtgtgtgg 60
 gcagcttcgc tttcggccgc ccgcgtggag aagctgtggg tgacgtgt ctacggcg 120
 cccgtgtgg aggaggccac caccacccct tttcggccca gcaacgcgc ggccatccac 180
 accggagggtc acaacgtgt ggccacccac gcttcgtgc ccaccgcaccc caacccccc 240
 gagatctgtc tggagaaac gaccgagaac ttcaacatgt ggaagaacaa catgtgtgg 300
 caagatcgac ggacatcat cagctgtgc gaccggggc tgaaggccctg ctgtggggcc 360
 cccctgtatca cccgccgtt cccaaggatgt atgttcgcgc ccatccccat ccactactgc 420
 gccccccgcg gttgcgcatt cttcggatgc aaacggcaaga atgttcaacgg cagggccccc 480
 tgcaccaac tgcgtggatgc gcgtgcacc cggcgtatcc ggccgggtgt gacgcaccc 540
 ctgtgtgtca acggcgatcc ggccggggg ggccgggtgt tgcgtggatgc aaacttcacc 600
 gacaacccca ggacatcat cttcggatgc aaggggggcc tggagatca ctgcacccgc 660
 cccaaacaaa acacccggca gatgcattacc atccggccccc ggccggccct ctacggccacc 720
 ggccgcatac tcggcgatcat ccggccggcc cactgcacaa tcacggccgc gaatgtggac 780
 aacacccatca aggacatgt gatccatgt caggccatgt tcggccacaa gaccatcg 840
 ttcaacggca gcacggccgc cgaccccgat atcgatgt acagcttcac ctggggggcc 900
 gagttttttt atcgcaacatc caccatgt ttcaacatgc cttcggacaa caccatcg 960
 cccaaacaaa ccaaacggcc catccatgt ccctggccca tcaacatgc catcaacccgc 1020
 tggcgggg tggggcaagg catgtatgc ccccccatcc ggggccatgt ccgtgtcgac 1080
 agcaacatca ccggccgtgt gtggatccgc ggccggcc aggatgtac caacacccac 1140
 gagatcttc ggccggccgc ccggccatgt ccggccatgt ggccggcc gctgttacaa 1200
 tacaagggtgg tggatgtca ggccctggggc gtggggccca ccaaaggccaa ggccggccgt 1260
 gtgcacggcg aggaggccgc cttcgcgtt ggccggatgt tccctgggtt ctctggggcc 1320
 gcccggccgc ccatggggcc ccgcgcgtt accctgtggcc tgcggccccc ccatgtgtcg 1380
 agccggatcg tcgtggatgc aaacacccatc tcgcggccca tgcggccca ccatgtgtcg 1440
 ctgtgtgtca ccgtgtgggg catcaacatgc ctcgaggccgc ggccgtgtggc ctgtggggcc 1500
 tacctgtgg accacccatgc gttggccatc tggggccatc ggccggcc gatctgtgcacc 1560

<210> 7
<211> 23
<212> DN
<213> 3m

<220>
<223> Description of Artificial Sequence: Leu122-Ser199

```
atccaggagc tgaagaacag cgccgtgagc ctgttgcacg ccatcgccat cgccgtggcc 2220
gaggggcaccc accgcatcat cgagggtgcc cagcgcatcg gccgcgcctt cttgcacate 2280
ccccgcgcga tccgcgcaggc ttgcgcagcg cccctgtgt aactcgacgc tgct 2334
```

<210> 8

<211> 2316

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Val120-Thr202

<400> 8

```
gaatccgcca ccatggatgc aatqaagaga gggctctgt gtgtgtgtgt gctgtgtgg 60
gcagtcgttcc ttgcgcggc cgccgtgag aagctgtgg tgaccgtgt acaaggcg 120
cccggtgtgg aaggaggccac caccacccctg ttctgcgcga ggcacgcggaa ggccctacgac 180
accggaggcg acaaccgtgt gcgcacccac gcctcgctgc ccacccgaccc caacccccc 240
gagatcggtgc tgagaaatcg gaccgagaac ttcaacatgt ggaagaacaa catgtggag 300
cagatcgacg aaggacatcat cagctgtgg gaccgacgc tgaaggctcg cgtggcgccg 360
gccacccagg ctgtcccaaa ggtgagttc gagccatcc ccatccacta ctgcgcggcc 420
gcccgggttcg ccatccctggc gtgcacgcgaa aagaatgtca acggcagcggg ccctcgac 480
aacgttgagca ctgttcgtgtg caccggccgc atccggcccg tggtagcgcac ccacgtgtcg 540
ctgaacgcga gcctggccga ggagggtgtg gtatgcgcga gggagaactt caccgacaac 600
gccaacgacca tcatgtgcg gtcggaaatcg aacgtggaga tcaactgcac ccggcccaac 660
aacaacaccc gcaagagatc caccatggc cccggccggc ctctctacgc caccggcgac 720
atcatcgccg acatccggc ggcggacttc aacatcgccg gggagaatgtg gaacaacaccc 780
ctgaagcaga tctgttgcacaa gtgcggacggc cagttcgca acaagacccat ctgttcaag 840
cagagcggcg ggccggaccc cgagatcggt atgcacatgt tcaactcggtt cggcgatgtc 900
ttctactatcg acacgcacca gctgttcaac agacatcggtt acaacacccat cggcccccac 960
aacaccaaccc gcaacatcatc ctgcgttgcggc cccatcgac agatcatcaaa cctgtggacg 1020
gggggtggca agggcatgtt cggccggccccc atccggccggc agatccgtcg cggcggacaa 1080
atcaccggcc tgcgtgtgcg cggcgacggc gggaaaggaa tcaagcaacac caccggatate 1140
ttccggcccg ggccggcgca catgcgcgcgaa aactggcgca gggatgtgtaa caagtacaag 1200
gtgggttggaa tgcggccatccc ggccgtggcc ccaccaacggg ccaaggccgg cctgtgtggcc 1260
cgcgagaacgc ggcgcgtgc acctggccggc atgttccgtt gcttccctggg cggccggccg 1320
agcaccatcg ggcggccggc ctgcacccctg acctggcgagg cccggccacgt gttgagggccg 1380
atcgtgcgcg aaggacaaatcg ctgtgtgcgc gccatcgccg cccagacgca cttgtgtgc 1440
ctgaccgtgtt ggggcatcaa gggatgtcg gcccggctgc tggccgtggaa gggatctctg 1500
aaggccggc acgtgtgtggg catctggggc tgcggcgca agtgcgtctg caccacccggc 1560
gtgccttggaa acgcacatgtt gggacaaacaaag agcttggaccc agatctgtggaa caacatgaccc 1620
tggatggagt gggagccggca gatcgacaaatc tacaccaacaa tgatctacac ctgtatgcgg 1680
gagagccggca acaggccggca gaaagaacggg cggagatgtc tggatgtggaa caatgtggcc 1740
agcttgcgttgc actgtgttgcg catcgacaaatcg tggatgtgggtt acatcaagat ctgcatcatg 1800
atcgtgtgggg gctctgggtggg ctgcgcacatc gttgcacccg tggatgtggat ctgtgtggcc 1860
gtggccggccgg gatcgacccgg cctgtgtttcc cagacccgtt cccggccccc cggccggccccc 1920
gacccggccgg agggatcgaa ggaggaggcc ggcggatcgcc acccgccggc cggccggccccc 1980
ctgggtgcacgc gctgtgtggc ctgtatgttgc gggccgttcgc gggccgtgtg cctgttccgt 2040
taccacccgcg tgcggccatctt gatctgtatc gccggccggc tggatgtggat gttgtggccg 2100
cgccgggtggg aggccctgttgc gatctggggc gggccgttcgc agatctgtatc ctgtgtggccg 2160
aagaacacccg cctgtgtggccatc gttgcacccg atccgcacatcg cccggccggc gggccggcc 2220
cgccatcgatcg aggtggccca ggcgcacccgc cccggcccttc tgcacatccc cccggccgtatcg 2280
ggccagggtt tggatgtggccg cctgtgtgtaa ctgcgtatcg 2316
```

<210> 9

<211> 2541

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Trp427-Gly431

<400> 9

```

gaattcggcca ccatggatgc aatagaagaga gggctctgtc gtgtgtgtgc gctgtgtgga 60
gcagtcttcg ttccggcccg cgccgtggag aagctgtggg tgaccgtgtc ctacggcggtg 120
cccggtgtggaa aggaggccac caccacccctg ttctcgccca ggcacgcacaa ggccctacgt 180
accggagggtgc acaacgtgtg ggcacccac gcctcgctgc ccacccgaccc caaccccccag 240
gagatcgatgc tgagaaatgtg gaccggaaatc ttcaacatgt ggaagaacaa catgtgtggg 300
cagatgcacg aggacatcat cagctgtgg gaccagagcc tgaagccctgc cgtgaagctg 360
accccccgtg gctgtgaccc tcaactgcacc aaccttgaaga acgcacccaa cacaaggagc 420
agaactgtggaa aggagatggaa ccgcggcgag alcaactgtt gcaagtttca ggtgtaccac 480
agcatcgccaa acaaaatgttggaa gaaggatgtt gccctgttca acaactgttgc cgtgtgtcccc 540
atcgacaaatcg acaacacccatcataaactgttgc acacccatgtt gatcaccacccg 600
gcctgtggccaa aggtgtggatcg cgagggccatc cccatcatgtt acatgtggccccc cgccgggttcc 660
gccccatctgtg aatgtcaacaa caaagaatgttca aacgcacccg gcccctgtgcac caacgtgtgg 720
acccgtgtggatcg gacccacccgg catccggccccc ttgtgtgtggaa cccacgtgttgc gctgtgtggc 780
agcctgtggccgg aggaggccgtt ggtgtatccgcg aacggaaatgttca tcaaccacacaa ccgcacccac 840
atcatcgatgc agtctgtggaa gaggtgtggatcg atcaactgttgc cccggcccccacaa caacaacacca 900
cgtcaagatgttca tcaatccgtt ccccgccggatcg gcttctatcg ccacccggatcg catcatgtgg 960
gacatccggccggaggccatctgttca acaatcatgttgc gggggaaatgtt ggaacaacaccc cctgtgtgg 1020
atcggtgtggccaa agtctgtggccgg ccatccgtggcc aacaaggatgtt gcaatgttca gcaaggatgttgc 1080
ggccggggaccggccatgttgc gatgcacacccgttca tcaactgttgcg ccggggatgttgc ttctatgttgc 1140
aacagccacccatgttgcgttca aacgcacccatgttgc gacccatccgttgc tggggcccccacaa caacacccac 1200
ggcaccatccatccgttgc cccatgttgc gacccatccgttgc tggggcccccacaa caacacccac 1260
atgtatggcccccgttgc gggccatccatgttgc gggccatccatgttgc gcaacatccatgttgc cggccgtgtgt 1320
ctgaccggccggccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1380
ggccggatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1440
cccccgtggccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1500
gtgaccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1560
cgccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1620
aacaacatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1680
atcaacatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1740
ctggggatcttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1800
agtcgtgtggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1860
cgccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1920
caggagaaga aacggatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 1980
ttcgcacatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2040
gtggggatcttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2100
agcccccgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2160
atcgaggaggaggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2220
ctggggatcttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2280
gacccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2340
ctgaaatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2400
agccgtgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2460
ggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2520
cgccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2580
cgccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc gggccatccatgttgc 2541

```

<210> 10

<211> 2541

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Arg426-Gly431

<400> 10

```

gaattcggcca ccatggatgc aatagaagaga gggctctgtc gtgtgtgtgc gctgtgtgga 60
gcagtcttcg ttccggcccg cgccgtggag aagctgtggg tgaccgtgtc ctacggcggtg 120
cccggtgtggaa aggaggccac caccacccctg ttctcgccca ggcacgcacaa ggccctacgt 180

```

accgaggatgc acaaactgtg gcccacccac gcctggcgtgc ccaccgaccc caaaccggcag 240
 gagatctgtc tgagaacgt gaccgagaa tcaacatgt ggaacacaa catgtggag 300
 cagatgcacg aggacatcat cagctgtgc gaccgagac tgaaggccctg cgtgaactg 360
 accccctgtg gctgacccct gcaactgcacc aacctgaga acgcccacca cacaagaggc 420
 agcaactgga aggagatgga ccggccgag atcaagaact gcagcttcaa ggtgaccacc 480
 agcatccgca acaaagatgca gaaggagatc gcccgttctt acaaagtgaa cgtggtrgccc 540
 atcgacacaacg acaaaccacg ctacaacatgc atcaactgtca acacccagctc gatcacccacg 600
 gctgtgcggca aggtgagctt cgaccccatc cccatccact acgtgcgcggc cgcgggtctt 660
 gccatctgtt agtgcaacgca caagaaggtc aacggcagcg gcccctgcac caacgtgagc 720
 accgtgcagt gaccccaacg catccggccc gtggtgagca cccagctgtg gctgaacggc 780
 agctggccgg agggaggcgtt ggtatcccgcc agggagactt tcaaccgaca cgccaaaggac 840
 atcgatctgtg agctgaaaggaa gacgtggatc atcaactgtca cccgccccca caacaacacc 900
 cggaaagagca tcacatccgg cccggccgc gccttcattgc ccacccggca catcatccgg 960
 gacatccggc accggccatc caactatcgcc gggcggagaat ggaacaaacac cctgaagcag 1020
 atcggtgacca agtgcggccg ccaggatggc aacaagatca tctgtttca gcaagacgc 1080
 ggccggccgc accggatgtt gatgcacagc tcaaccatggc gccggcggatgt ctcttactgg 1140
 aacaggaccc accgttcttca cagcaaccttca aacaaccatca tccggccccca caacaccac 1200
 ggcacccatca ccctggccctt ccggatccatcg cagatcatca accggccggg cggcaaggcc 1260
 atgtacggcc accgttccatcg cggccatggc cgtgtccggca gcaacatcatc cggctgtctg 1320
 ctgaccggcg acggccggca gggatcatgcg aacaccacgg agatcttcgg cccggccggc 1380
 ggccgatgcg gggacaactcg ggcggccatcg ctgttcaactg acaaagggtgtt gaaagatcgac 1440
 cccctggggcc tggcccccac caaggccaa gccggccgtt tgccggccgc gaaaggccggc 1500
 gtgacccttgc ggcgcattttt cctgggttcc tggggccggc ccggcagcac catggggccgg 1560
 cggccatgcg ccctggccatcg gggccatgttcc cggccatgttgc gggccatgttgc gacggccggc 1620
 aacaacccatcg tggccggccatcg cggccggccatcg cagcaacttcg tgccgttgc cgtgtggggc 1680
 atcaagacgc tgccggccgc ctgtgtccggc gtggagccgtt acctgtggaa ccacggatgt 1740
 ctggccatctt gggggatcgcc gggcaactcg attcgcacca ccggccgttgc ctggaaacggc 1800
 agtggagacca acaaaagatcc gggccatgttgc tggaaacacaatc tgacttggat ggagtgggg 1860
 cggccatgcg aacaactacat ccaactgttgc tccaccatgttgc tgccggaggag ccagaacacc 1920
 cggccatgcg acaactatcg cggccatgttgc tgccggatgttgc tggccatgttgc tgccggatgttgc 1980
 ttcgcacatcg gggccatgttgc tgccggatgttgc aatgttca tcatgtatgttgc tgccggccgtt 2040
 gtggccatgc gcatgtgttgc caccgtgttgc acatgttgc accggccgttgc ccagggttgc 2100
 agccccctgtt gtttccatcg ccgcctttccccc gccccccggc gccccggccgg cccggggggc 2160
 atcgaggagg tggccggccgc ggcggccatcg gaccggccggc gccccctgttgc acggccgttgc 2220
 ctggccatgttgc ttttggccatcg ctttgcggccatcg ctgtgttgc tgacttccca ccggccgttgc 2280
 gaccgttgc ttttgcggccatcg ctttgcggccatcg tgacttccca ccggccgttgc 2340
 ctgttgcacttgc ttttgcggccatcg ctttgcggccatcg tgacttccca ccggccgttgc 2400
 agctgttgc acggccatcg catccggccatcg gccggccatcg acggccatcg catccggccatcg 2460
 gcccacccatcg ttttgcggccatcg gccggccatcg acggccatcg catccggccatcg 2520
 cggccatcg ttttgcggccatcg gccggccatcg gccggccatcg acggccatcg catccggccatcg 2541

<210> 11

<211> 2541

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Arg426-Gly431B

<400> 11

gaaatccgcca ccatggatgc aatggaaagaga gggctgtgttgc tggtgtgttgc gttgtgttgc 60
 gcaatccatcg ttccggccatcg cggccgttgc aaggctgtgg tgaccgttgc ctacgggttgc 120
 cccctgttgc gggaggccatcg caccacccatcg ttccggccatcg gggccatcg gggccatcg 180
 accggatgttgc acaacatgttgc gggccatcg gggccatcg ccaccggccatcg caaaccggccatcg 240
 gagatctgttgc tggggaaatcg gaccggaaatcg ttcaacatgttgc gggccatcg gggccatcg 300
 cagatgcacg aggacatcat caccgttgc gaccggaaatcg ttcaacatgttgc gggccatcg 360
 accccctgttgc gggccatcg gggccatcg aacccatgttgc gggccatcg gggccatcg 420
 agcaactgttgc agggatgttgc cccggccatcg atcaactgttgc gggccatcg gggccatcg 480
 agcatccgcca acaaaggatgttgc gggccatcg gggccatcg gggccatcg gggccatcg 540

atccacaacg acaacacccg ctacaactgc atcaactgcg acaccagegt gatcaccccg 600
ggctggccca aggtgaggctt cgccggccat ccacccatcc actggcccccc cggccggccctt 660
gcacatcttgc agtgcggca caaaggctt aacccatcc actggcccccc cggccggccctt 720
accgtgtcagt gacccaccacg catccggcccc aacccatcc actggcccccc cggccggccctt 780
agccgtggccg aggaggggccgt gtgttgcacccg aacccatcc actggcccccc cggccggccctt 840
atccatctgtc agtgcggaa gacgttggag atccatccactgc cccggccccaa caaacacacc 900
cgccaaagacg tccatccatgg cccggccggcc ggtttccatcc actggcccccc cggccggccctt 960
gacatcccccc agggccctttcgtt ccaatccatgc gggccggaaat gaaacacacc cttccatccactgc 1020
atccgtgtccca agtgcggccg ccaggatccgc aacccatccactgc gggccggaaat gaaacacacc cttccatccactgc 1080
ggccggccgacc cccatccatgc gatgcacccg aacccatccactgc gggccggaaat gaaacacacc cttccatccactgc 1140
aacacccacc cccatccatgc tccatccatgc aacccatccactgc gggccggaaat gaaacacacc cttccatccactgc 1200
ggccatccatca cccatccatgc tccatccatgc aacccatccactgc gggccggaaat gaaacacacc cttccatccactgc 1260
atgtacggccc cccatccatgc tccatccatgc aacccatccactgc gggccggaaat gaaacacacc cttccatccactgc 1320
ctgaccggcg acggccggca gggatccatgc aacccatccactgc gggccggaaat gaaacacacc cttccatccactgc 1380
ggccatccatgc gggatccatgc gggccggca gggatccatgc tccatccatgc aacccatccactgc gggccggaaat gaaacacacc cttccatccactgc 1440
ccccctggggcc tggcccccacg caaaggccatgc cccatccatgc tccatccatgc aacccatccactgc gggccggaaat gaaacacacc cttccatccactgc 1500
gtggccatcccg gggccatccatgc tccatccatgc aacccatccactgc gggccggca gggccggccatgc tccatccatgc aacccatccactgc 1560
cgccatccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc 1620
aacacccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc 1680
atccacaacg tccatccatgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 1740
ctggggatccatgc tggggccatgc cggccggccatgc atccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 1800
agctggggatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 1860
cgccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 1920
caggagggatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 1980
ttccatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2040
gtggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2100
agccccccatgc gtttccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2160
atccggggatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2220
ctggggatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2280
gaccatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2340
gtggggatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2400
agccatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2460
ggccatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2520
cgccatccatgc tggggccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc aacccatccactgc gggccggccatgc tccatccatgc 2580

<210> 12

<211> 2541

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Arg426-Lys432

<400> 12

gaatttcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt gctgtgtgg 6
gcaatgtttt ttccggcccg ccggccggaa aagctgtggg tgaccgtgtta ctacggcggt 120
cccgctgtggaa agggacccggc caccacccgtt ttcggccca ggcacggccggatggccgtacg 180
accggatgttcc aacaatgtgttgg gggccacccggaa gcttgcgtgc ccacccggacc caaacccggatc 240
gatggatgttgc tgaaatggatgttgg aacccggaaatgttccatgttggaaatgttggaaatgttgg 300
cagatgcacgg agacatcatatgcgactgttgg gaccggacggccatgttggaaatgttggaaatgttgg 360
accccccgttggtggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 420
ggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 480
ggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 540
ggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 600
ggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 660
ggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 720
ggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 780
ggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 840
ggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttggatgttgg 900

cgaaagagca tcaccatcg cccggggcgc gccttctacg ccacggcgca catcatcgcc 960
 gacatccgcg aggccactg caacatcagc ggccgagaatg ggaaacacac cctgaagcgcg 1020
 atcggtgacca agtcgcggc ccagtgtccg aacaagacca tctgtttcaag gcagagcgcg 1080
 gggggcggacc ccggatctgtg gatgcacacgc ttcaactcgcc ggccggatgtt ttctactgcg 1140
 aacagcaccc agtcgttca a cggacacttgg aacaacacca tcggccccaa caacaccaac 1200
 ggacccatca cccctggccctg ccgcacatcaag cagatcatca accggcgccgg caacaagggc 1260
 atgtacgccc ccccccattcg ccggccagatc cgctgcagca gcaacatcgc cggctctcg 1320
 ctgaccccgcg aegggcgccaa ggagatcagc aacaccacccg agatcttccg cccggcgccg 1380
 ggcgcacatc ggcgacaaactg ggcgagcgag ctgtacaaatg acaaggltgtt gaagatcgcg 1440
 cccctggggc tgcccccacca a cggccaaatg cggccgttgg tgccgcgcgaa gaagcgccg 1500
 gtgacccctgg ggcgatctgtt cctgggttccg cggccggccg cggccgcacatc catggccgcg 1560
 cgcacgcgtg cccctggccctg ccgcacatcgcc cagtcgttca gggccatctgt gcaagcgcg 1620
 aacaacctgc tgccgcgcatc egaggcccgac caggccatcg tgca gctgcg cgtgtggggc 1680
 atcaacgcg ctcggcccgcc cgtgtggcc gttggagcgatc acctgaaaggcc caagcagctg 1740
 ctggccatot gggggctcgcc ggccaaactg atcgacca cccgcgtggcc ctggaaacccg 1800
 agctggggcgcg aacagacccatc ggacccatcgcc tggaaacacca tgaacctggat ggagtggggc 1860
 cgcgcacatc acaactacac ccacactgatc tacacccttgc tgccggaggag ccagaacccg 1920
 caggagaaaga acgagccggc gtcgttggcc ctggccaaatg gggccacccgtt gtggaaactgg 1980
 ttgcacatca gcaatgtgtt gtgttacatc aagatcttgc tcatgtatcg tggccggctc 2040
 gtggccctgt gcatctgtt caccctgttccg aacatcgta accggcgatcgcc ccagggttac 2100
 agcccccgtt gcttcccgac ccgcattccccc gcccccccgcc gccccccggcc ccccgaggcc 2160
 atcgaggaggaggccggccgcg ggcgcgcacgc gacccgcaccc gcccccttgcgtt gcacccgttgc 2220
 ctggccctgtt ctggggacgcg ctcgcgcacgc ctgtgttgcgtt tcaacttacca cccgcgttgc 2280
 gacccgtatcc tgcgcgcgcg cccatcgatc gatgttgcgtt gccggccggcc ctggggaggccg 2340
 ctgaaatgtt gggggccaaactt gtcgcgtatc tgatccagg agtgcgaaaga cgcgcgttgc 2400
 agccgttccg acgcacatcgcc catcgccgtg cccggggccg cccacccatc catcgatgg 2460
 gcccacgcgc tccggccgcgc ctccctgcac atccccccgcg gcatccgcgc gggcttgcgt 2520
 cgccgcctgc tgtaactcgatc g 2541

<210> 13

<211> 2535

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Asn425-Lys432

<400> 13

gaattccgcga ccatggatgc aatgaagaga gggtctctgt gtgtgtgtgt gctgtgtgg 60
 gcacgttccg tttggcccg cccgttggag aacgtgtggg tgacccgttgc ctacggcgctg 120
 cccgtgttgcg aggaggccac caccacccctt ttcgcgcac gcaacgcacca ggcttacac 180
 acccgaggatgc aacacgttgc ggcacccatc gcttgcgcgc ccacccgacca caacccccc 240
 gagatcgatc tggaaacatcg gaccggaaatc ttcaacatgtt ggaacaaatc catgttggag 300
 cagatcgacccg acacatcatc cagctgttgc gaccggatgc tgaacccctgg cgttgaaatcg 360
 accccctgtt ggtgttgcctt gcatcgacc gacccatcgaa aacccatcgaa acgcacccaa caccacggc 420
 agcaacttgcg aggagatggc cccggccggatc atcaaaatcg gcaatcgatc ggtgaccacc 480
 agcatccgcg acaaatgtgc gaaaggatgc gccccttttcc tcaacgttgc cgtgtgtggcc 540
 atcgacaaatcg aacacccatcg ttcaacatcg tcaactatcg acaccacgtt gatccccc 600
 gccctggccca aggttgcgtt ccgcacccatc ccacatccatc actgcgcggcc cccggccgttcc 660
 gccatcttgcg atgttgcacca caaaggatcc aacggccggcc gccccctgcgc caacgttgacg 720
 accgtgcgtt gacccacccg catcccccgtt gtgttgcgtt cccatcgatc gtcgttgcgtt 780
 agcttggccg aggaggccatc ggtgtatccgc acggatgcactt tccacgcacca cgcacccatcg 840
 atcatcgatc agctgttgcg aacatcgatc gacccatcgaa caacacccatcg caacacccatcg 900
 cgcacccatcg tccacatcgcc ccccccggccgc gcccctttccg ccacccggccg catcatcgcc 960
 gacatccgcg aggccactgc caacatcgatc ggcacccatcg gcaacacccatcg cccggccgttcc 1020
 atcggttgcacca agctgttgcgc ccaggatctgtt gatgcacacgc ttcaactcgcc gggccggatgtt ccgttactcg 1140
 aacagcaccc agtcgttca a cggacacttgg aacaacacca tcggccccaa caacaccaac 1200
 ggacccatca cccctggccctg ccgcacatcaag cagatcatca acgcacccaa ggcacatcgatc 1260

ccccccccca tcgcggcca gatccgtgc agcagaaca tcaaccggct gctgtgtacc 1320
 cgcgacggcg gcaaggagat cagcaacacc accggagatct tccggccccc cgccggcgac 1380
 atgcgcgaca atgcgcgac cgatccgtac aagtcaacgg ttgttgaatgt cgagccccct 1440
 ggccgtggcc caacaaacggc caageggcgc gtggtgacgc gcgagaaacg cgccgtgacc 1500
 ctggggccca ttgttccctggg ttccctgggc ccggccggca gcaccatggg cgccggcagc 1560
 ctgcggccca ccgtggcggc ccgcacgtgc ctgagccggca tcgtgcacga gcagaacaca 1620
 ctgtgcgcg ccatacgaggc ccacgcacac ctgtgcacgc tgaccgtgtg gggcatcaag 1680
 caagctgcagg ccgcgcgtgt ggccgtggag cgctacccgta aggaccaacga gtgtgtggc 1740
 atctcgccgt gcagccggca gctgtatccgc accaccgcgc tgccctggaa cgccagctgg 1800
 agcaacaaga gctgtggaca gatctggaa acatcatgtt ggatggatgt ggacgcggag 1860
 atgcgcgaaact acaccaactt gatccgtgggg agagccggaa ccacgcggag 1920
 aagaacggacg agaggactgtt ggatgtggac aagtggggca gctgtggaa ctgttgcac 1980
 atcagcaagt ggctgtggta catcaagatc ttcatcatga tcgtggccgg ectgtgtggc 2040
 ctgtgcacgtc ttgttccatgt gtcgacatc gtgaacccgc tgccggccgg ctacacccc 2100
 cttagcttcc agacccgggtt ccccccggcc ccggccggcc acggccccca gggcatcgag 2160
 ggaggggggcc gcaaggccggc ccgcgcggcc acggcccccc ttgttgcacgg ctgtgtggc 2220
 ctgtatctggg acacacttggc cgcctgtgc ctgttgcaccc accaccggct gecgcacgtt 2280
 atccgtatcg ccgcggccat ctgtggatgt ctggggccgg ggggttggga gggctgtgaag 2340
 tactggggca acatctgtca gtactggatc caggagctga aaaaacgcgc cgtgagccgt 2400
 ttgcacggca tcgcacatcgc ctgtggccgg ggcacccggc gcatcatcgaa ggtggcccg 2460
 cgcacatggcc ggcgccttcc gcacatcccc ccgcgcacatcc gecagggtt cgcgcgcgc 2520
 ctgtgtgttac tcgac 2535

<210> 14

<211> 2529

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Ile424-Ala433

<400> 14

gaattcggcca ccatggatgc aatgaagaga gggctctgtc gtgtgtgtgtc gctgtgtggaa 60
 gcacgttccg ttccggccgg cgcggccggaa aactgttggg tgacccgtgtc ctacggccgt 120
 cccctgtggaa aggaggccac cacccatctg ttccggccca ggcacgcacaa ggccatcaca 180
 accggagggtc acaacgtgtg ggccacccccc gcttcggcgc ccacccggcc caacccccc 240
 gagatctgtcg tggagaacat gccggaaatc ttcaatcatgt ggaagaaacaa catgttggag 300
 cagatgcacg agacatcatc cgcgttgc gaccagatgc tgaaggccctg cgtgtggctg 360
 acccccccgtt ggcacgttgc gacatgcaccc aacccgttgc aacgcaccaaa caccaagggc 420
 agcaacttggaa aggagatggaa ccggccggag atcaactatgc gcaatccatca ggtggaccacc 480
 agcatccgc acaagatgcg gaaggaggatc gcccgttctt acaactgttgc cgtgtgtccc 540
 atcgacaaacg acaacaccgg ctacaactgtc atcaactgtca acaccaggctt gatccaccccg 600
 gcccgtccccca aggtggatctt cgcggccatc cccatccatct actgcgcggcc cgccgggttc 660
 gccatctgtcc agtgcacggaa caaaggaaatc aacccggccgg gcccctgcac caacatgttgc 720
 acccggtcgttgc gacccacccgg catccggccccc qtgggttgcacca ccaacgttgcgttgc 780
 agccgtggccg aggaggccgtt ggtgtatccgc agcggaaactt tcaccgcacaa cgccaaagacc 840
 atcgatctgtc agtgcacggaa gaggtggggatc atcaactgtca cccggccccc caacaaacacc 900
 cgcacatccgcg tccatcatcg ccccccggccgg cccttcttgcg ccacccggccgca catcatcgcc 960
 gacatccgcg aggcccaactt ccaatcatcgcc ggcggaaatgt ggaacaaacac cctgttgcac 1020
 atcgatgttgcgaa agtgcacggcc cccatccgttgc aacaaggatca tcgtgttgcac gcaagacgc 1080
 ggccggccacc cccatccgttgc gatgcacgc tcaactgtcg gggccgtgtt ctctactgtc 1140
 aacatccgcg agtgcacggaa cccatccgttgc aacaaggatca tcgtgttgcac gcaacccacc 1200
 ggccatccatcg cccatccgttgc gatgcacgc tcaactgtcg gggccgtgtt ctctactgtc 1260
 cccatccgcg gccatccatcg cccatccgttgc gatgcacgc tcaactgtcg gggccgtgtt ctctactgtc 1320
 ggccggccacc gatgcacggaa cccatccgttgc gatgcacgc tcaactgtcg gggccgtgtt ctctactgtc 1380
 gacaactgtcc gacgcgcgtt gatcaactgtc aagggttgcac gatgcacgc tcaactgtcg gggccgtgtt ctctactgtc 1440
 gcccccccacc gatgcacggaa cccatccgttgc gatgcacgc tcaactgtcg gggccgtgtt ctctactgtc 1500
 gccatctgtcc tggggccccc ggcacatccgc tcaactgtcg gggccgtgtt ctctactgtc 1560
 ctgcacccgttgc agggccggccaa gatgcacgttgc ggcacatccgc tcaactgtcg gggccgtgtt ctctactgtc 1620

cgcgccatcg aggcccacga gcacactgtc cagctgaccg tttggggcat caagcagctg 1680
 cagggcccgcg tgcgtggccgt ggagcgatc ctgaaggaccg agcagctgtc gggcatctgg 1740
 ggtcgacggc gcaacgttgc ctgcacccatc ggcgtgcctt ggaacgcacg ctggagcaac 1800
 aagagcttgc accatgtatc gaacaacatc acctggatgg agtggggacgc cgagatcgac 1860
 aactacatcca acctgtatca caccctgtc gaggagagcc agaaacccaga ggagaagaaac 1920
 gagcaggaggc tgcgtggatgt ggacaatggg ggcagccctgt ggaactgttgc cgacatcgac 1980
 aagtggctgttgc ggtatcatcaaa gatcttcattc atgtatgttgc ggggttgcgtt gggctcgcc 2040
 atcgtgttca cctgtgtcgatc catgtgtaaac cgcgtgcggcc agggctacatc cccctctgagc 2100
 ttccagaccc gtttcccccgc ccccccggcc cccgacccgc cccgtgttgc acggccctgtt ggcctgtatc 2160
 ggccggcggacg cgcacccggaa cgcacggcgc cccctgttgc acggccctgtt ggcctgtatc 2220
 tggggacggacc tggcgacggcgtt gtgcgttgc acgttacaccatc gcttgcgcga cctgtatcgcc 2280
 atcggccggcc gatcatgttgc gtcgtggcc ggcggccgtt gggaggccctt gaagacttgcg 2340
 gcaacccgtc tgcaatgttgc gatccaggag ctgaaagaaca ggcgcgttgc gctgttgcg 2400
 gccatcgccca tccgtgtcccg cggggccacc gaccgcatac tcgagggttgc ccagccgtatc 2460
 ggccggccgtt tctgttgcacatc ccccccggcc atccggccagg gtttgcgttgc cccctctgttgc 2520
 taactcgag 2529

<210> 15

<211> 2523

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Ile423-Met434

<400> 15

gaattcgccca ccatggatgc aatgaagaga gggctctgtt gtgtgtgtgtc gctgtgttgc 60
 gcagcttcgg tttcgcccgac cgcgtggag aagctgtggg tgaccgtgttca ctacggcggt 120
 cccgtgtggg aggaggccgc caccacccctt ggcgtgcggca ggcacgcacaa ggcccttaacgac 180
 accgggtgtc acaacatgttgc ggcacccacac ggcgtgtccgc ccacccggcc caaccccccac 240
 gagatgtgtc tgagaaatgttgc gaccggatgttca acatgttgc ggaagaaacaa catgtgtgg 300
 cagatgtcaccg agacatgttgc acgtgtgttgc gaccggatgttgc tgaagccctgtt cgtgtggatgt 360
 acccccccgttgc gctgtgttgc gactgttgc accatgttgc gacccacccaa caccacaaagc 420
 agcaactgttgc agggatgttgc cccggccggatc atcaaaatgttgc gcaactttca ggtgaccaccc 480
 agcatcgccca acaaaatgttgc gaaaggatgttgc gcccttttgc acatgttgc gttgtgttgc 540
 atcgcacaaacg acacccacccatc acatgttgc acacccatgttgc gatccacccatc 600
 gctgtgttgc acgtgtgttgc cggccggccatc cccatgttgc acatgttgc cccggccgttgc 660
 gccatctgttgc agtgcacaaatc caaaggatgttgc aacggccatc ggcctgttgc acacgttgc 720
 accgtgttgcgttgc gacccacccatc cccggccggatc gtcgtgttgc gccatgttgc gctgtgttgc 780
 agcgtgttgcgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc cccggccgttgc 840
 atcgtgttgc acgtgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc cccggccgttgc 900
 cgcacccatc cccatgttgc acatgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 960
 gacatcgccca agccggccatc cccatgttgc acatgttgc gggggccgttgc ggtgtgttgc acggatgttgc 1020
 atcgtgttgcgttgc acgtgtgttgc cccatgttgc acatgttgc gggggccgttgc ggtgtgttgc acggatgttgc 1080
 ggccggccatc cccatgttgc acatgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1140
 acacccatc cccatgttgc acatgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1200
 ggccggccatc cccatgttgc acatgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1260
 cggccggccatc cccatgttgc acatgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1320
 aaggatgttgc acatgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1380
 tggcgccatc acgtgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1440
 accaaggccatc acgtgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1500
 ttctgttgcgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1560
 gtgtgttgcgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1620
 atcgtgttgcgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1680
 cggccggccatc cccatgttgc acatgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1740
 agcggccatc acgtgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1800
 ctgtgttgcgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1860
 accaaggccatc acgtgttgc gggggccgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1920
 gagatgttgcgttgc ggtgtgttgc acggatgttgc tcaatgttgc 1980

ctgtggtaca tcaagatctt catcatgatc gtggggggcc tggggggctt ggcacatcg 2040
 ctacccgtgc tgagcatgtg gaacccgtg cggccagggtt acacccccctt gagttccatg 2100
 accccgttcc cccgcggcccg cggccggcccg cggccggccgg gcatcgaggaa ggaggccggc 2160
 gagcgcgacc gcgaccggca cggccggccgg gcatcgaccgc tgctggccctt gatctgggac 2220
 gacccgtgcga gctctgtctt gttcaactac cggccgtcg ggcacccgtt gctgtatccg 2280
 gcccgcattcg tgtagctgtt gggccggcccg ggctgggggg cccctgaatgtt ctggggcaac 2340
 ctgtcgactt actggatcca ggactgttgg aacacggccgg tgagctgtt cggccatccat 2400
 gccatcgccg tgcccgagggg caccggccgg atcatcgagg tgcccgaggc catcgccgc 2460
 gcccgttccgc acatcccccc cccatcccg cagggttccg agcgcgcctt gctgttaactt 2520
 gag 2523

<210> 16
 <211> 2517
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Gln422-Tyr435

<400> 16

gaattccgcca ccattggatgc aatagaagaga gggctctgtt gttgtctgtt gctgttgga 60
 gcacgtttcg ttggcccgcc cgccgttgag aagctgtggg tgaccgtgttta ctacggctgt 120
 cccgtgttga aggaggccac caccacccgtt ttctgtccca ggcacgcacca ggcctacgac 180
 acccgatgtc acacatgttgg ggcacccac ggcctgtgtt ccacccgaccc caaccccccac 240
 gagatgttgc tgtagactgttgg gaccggaaatc ttcaatctgtt ggaagaaacaa catgttgaa 300
 cagatgttgc acatgttgcggccatc cggccgttggg gacccaggcc gttaaaggccctt cgtgtaaatgtt 360
 accccctgttgg ctgtgttgcggccatc gacccgttgggg aacccgttgggg acggccacca caccatgttgc 420
 agcaacttggg aagtagatggg ccggccggag atcaaaactt gcaatgttcaaa ggttggccacc 480
 agcatccgcg acacatgttgg gaaaggatgttgg ggcctgtgtt acacatgttgg gttggggccccc 540
 atcgatccgcg acacatccggccatc acatgttgcggccatc gacccgttgc gatccacccatc 600
 gctgttgcggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 660
 gccatctggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 720
 acccgatgttgc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 780
 agccgttgcggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 840
 atcgatccgcg acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 900
 cggaaagacatc tccatccgttggccggccatc acatgttgcggccatc acatgttgcggccatc 960
 gacatccgcg acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1020
 atcgatccgcg acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1080
 ggcggccggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1140
 aacaggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1200
 ggcacccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1260
 cagatccgcg acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1320
 atcgatccgcg acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1380
 agccgttgcggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1440
 gccaaacatgttgcggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1500
 ggtttctggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1560
 ggcggccggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1620
 gcccggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1680
 ctggccgttgcggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1740
 aacaggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1800
 cagatccgcg acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1860
 ctggatccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1920
 ctggatccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 1980
 tacatccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 2040
 ttccctggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 2100
 gacccggccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 2160
 ctggatccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 2220
 ctggatccatc acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 2280
 atcgatccgcg acatgttgcggccatc acatgttgcggccatc acatgttgcggccatc 2340

```

cagactggaa tcaggagct gaagaacgcg ggccgtggcc ttgtcgacgc catcgccatc 2400
ggccgtggcc agggcaccga cccatcatc gagggtggcc agccgatcg cgccgccttc 2460
ctgcacatcc cccgcgcata cccgcaggc ttccgagcgcc cccctgtgtaa actcgag 2517

<210> 17
<211> 2517
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Gln422-Tyr435B

<400> 17
gaattcgccca ccatggatgc aatggaaagg gggctctgtgtgtgtgtgtgtggaa 60
gcagtttcgtt ttccggcccg cgccgtggaa aggtttgtgg tgacccgtgtatccatcgatcgt 120
ccctgtggaa aggaggccacg caccatccgtt ttcgtggccaa ggccggccaa ggccgtacac 180
accggaggatc aacaatgtgtt ggccaccccaat gctcgctgg ccacccggccaa caccggccatc 240
gatgtatgttgc tggaaaggatc gggcggaaac ttcacatgtt ggaaggaaacaa catgggtggaa 300
cagatgcacg aggacatcat cggctgtgg cggacacggcc tgaaggccgtt cttgtggatgtt 360
acccgggtttt ggtgtggccctt ggtatgtttttt aactttgtttt aactttgttttccatcgatcgt 420
aacaatgttggg aaggatggcc cggccggccggatc aacaaatgttggccatcgatcgttccatcgatcgt 480
agatcccgaa cacaatgttgg aaaaaggatggcc tggccgtttttt aacaatgttggccatcgatcgttccatcgatcgt 540
atcgatcgttgg aacaacccggatc ttacatgttgc atcaatgttgc aacccatgttgcgttccatcgatcgt 600
gcctgtggccca aggttggatgtt ccggccatccatc cccatccatgttgc atctggccccc cggccgttccatcgatcgt 660
gcattatgttgc atgttgcacgg acaatggatcggcc aacccggccggatccatcgatcgttccatcgatcgt 720
acccgttgcgttgc aacccggccggatccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 780
aactctggcccg aggagggggtt ggtgtatccggc agggatggatc aacccatgttgcgttccatcgatcgt 840
atcatcgatcgttgc atgttgcggatc gggctgtggatc atcaatgttgc cccggccggatccatcgatcgt 900
ccaaatgttggccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 960
gacatccggccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1020
atcgatgttgc aactgttgcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1080
ggccggccggccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1140
aacaatgttggccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1200
ggccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1260
cagatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1320
atcgatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1380
aagccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1440
ggccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1500
ggcttcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1560
ggccggccggatccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1620
ggccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1680
ctggccgtgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1740
aaatgttgcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1800
cagatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1860
ctgtatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1920
ctggatgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 1980
tataatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 2040
gtgtcgatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 2100
ttcccgccccc cggccggccccc cggccggccccc gaggccatcgatcgttccatcgatcgttccatcgatcgt 2160
gaccggccggatccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 2220
cgccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 2280
atcgatgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 2340
cagatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 2400
ggccgtggccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgttccatcgatcgt 2460
ctgcacatcc cccgcgcata cccgcaggc ttccgagcgcc cccctgtgtaa actcgag 2517

```

<210> 18
<211> 2322
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Leu122-Ser199;
Arg426-Gly431

<400> 18

gaattcggcca ccatggatgc aatgaagaga gggctctgtc gtgtgtctgtc gctgtgtgga 60
 gcagtcttcg ttttgcggag cgccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctggcca ggcacgcaca ggcctacgca 180
 accggagggtc acaacgtgtc ggccacccac gctgtgtgc ccaccgaccc caaccccccag 240
 gagatctgtc tggaagacgt gaccggaaac ttcacatgtt ggaagacaca catgtgtggag 300
 cagatcgacg aggacatcatc cagctgtgg gaccggaaqcc tgaaggccctg cgtgaagctg 360
 ggcacacagcg tttatcaccata ggcctgtccc aagggtgtc tgagccat cccccatccac 420
 tactgcggcc cccgcgggtt ccgcacccctg aatgtcaacg acaagaaggtt caacggcagc 480
 ggcccttcggc acaacgtgtc gaccgtgtgg tgccacccacg gcatccccc cgtgtgtgago 540
 aacccagctgtc tgctgttacggc cagctgtggc gaggaggccg tgggtatccg cagcggaaac 600
 ttcaacggaca acggcaacacg catcattgtc cagctgtgg aagacgtggaa gatcaactgc 660
 aacccggccccc acaacaaacacg cggcaagacg atacacatgg gcccccggcc cgccttctac 720
 gccacccggcg acatcatcgcc gcaatcccg caggccccact gcaacatcg cggcgagaa 780
 tggaaacaaca cccgtaaacgca gatcgtggac aagctgtggg cccagggttggg caacaagac 840
 atccgtgttca agacggacagc gccggcgac ccggatgttca acagcacctg gaacaacacc 900
 ggccggcgagt ttcttactgtt caacacgacc cagctgttca acagcacctg gaacaacacc 960
 atccggccggcc aacacacacca cggccacccatc accctgtccctt gccggatcaa gcaatctc 1020
 aacccggccggc gggcgaaacggc catgtacgc ccccccattt gccggccatgg cccgtcgacg 1080
 agcaacatccg cccgtgtgtc gtcggacccgc gacggggcga aggagatcgca caacacccac 1140
 gagatcttcg ccccccggcg cggccgtatcg cccgcacatc ggcgcacatcg gctgtacaa 1200
 tacaagggtgg tgaatgtca gcccctggcc gttggccccc ccaaggccaa ggcggcgctgt 1260
 gtgcagccgcg aacggccacgc cgtggaccctt ggcggccatgtt tccgtgggtt cctggggccgc 1320
 gccggccggca cccatggggcc cccgcacccgtt accctgtggc tgccggccgg ccacgtcgctg 1380
 aqccggccatcg tccgtacgcga gaaacaaacctt ctggggccca tccggggccca gaaacccatcg 1440
 ctgcgtgtc cccgtgtgggg catcaacggcgtc ctcggggcc ggcgtgtggc cgtggagccg 1500
 tacctgttggg acccggatgtc gtcggccatc tgggggtgtca gggccaaactt gtcgttcacc 1560
 accggccgtgc cccgtaaacgc cccgtgtggc aacaaaggccg tggaccatgt ctggaaacaa 1620
 atggacctgtc tggtggatgggg ggcgcacatcg gacaaactt ccaacccatgt ctacaccctg 1680
 atccggggaga gccggaaacca gcaacggaaacg aacggacgtgg agctgtgtgg gctggacaa 1740
 tggggccggcc tttgttggacttc gttccgtatcg acaactgtgg tgggtatcatc caagatcttc 1800
 atccatgtatcg tggggccggcc ggtggggctt cgcacatgtt tcaccgtgtc gacgtatcg 1860
 aacccgtgtc gccaggggccatc cccggccctgtc aggttttttttccggatccca gggcccccggc 1920
 ggcggccggcc gcccccggggg catcgaggag gggggggggc agggggggccg ccacccggccg 1980
 agcccccctgg tccacggccgtc gtcggccctt atccggggacg acctggccgcg cctgtgtccgt 2040
 ttccgtatcc accggccgtccg cccactgtatcg tgcacccatcg cccgcacatcg ggagctgtgtc 2100
 ggcgcggccggcgttggggggcccttggaaactt tggggaaactt tggggatgttccatc tggatccatc 2160
 gagctgtggaa acggccggccgttggggatgttccatc tggggatgttccatc tggatccatc 2220
 accggccggccatc tccatggggccggccggccatcg cccatccggccg cccatccggccg 2280
 cgcacatccggccatc tccatggggccggccatcg ggcggccctgtc tggtaactcg ag 2322

<210> 19

<211> 2322

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Leu122-Ser199;
Arg426-Lys432

<400> 19

gaattcggcca ccatggatgc aatgaagaga gggctctgtc gtgtgtctgtc gctgtgtgga 60
 gcagtcttcg ttttgcggag cgccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120

cccgtgtggaa aggaggccac caccacccctt ttctggccca ggcgcgcacaa ggcgtacgac 180
 accggaggtc acaacgtgtg ggcacccac gctctggcgc ccacccgaccc caacccccc 240
 gagatctgtc tgagagaacgt gaccggaaat ttcacatgtg gaaagacaa catgtgtggaa 300
 cagatgcacg aggacatcatcag ctggctgtgg gaccaggaccc tgaaggccctg cgtgaagctg 360
 gcaacacgcg tgcataccca ggcctgcccc aaggtgagct tcgagccat ccccatccac 420
 tactgtggcc cccgcgcgtt cgcacatctgtg aagtgcacaaag acaagaagtt caacggcagc 480
 ggccttgcac caacacgtgac caccgtgcac tgacccccc gatccggcccc cgtgttgagc 540
 accccaggctc tgctgaaacgg cagcctggcc gaggaggccg ttgtgtatccg cagcggaaac 600
 ttccaccgaca acgcccacac catcatgtg cagctgaaagg agagctgtgg gatcaactgc 660
 accccggccca aacaacaacac ccccaacagac atccacatc gccccccggcc cgccttctac 720
 gccacccggccg acatcatcgcc cgcacatccgc gggcccccact gcaacatcg cggcggaaag 780
 tggaaacacaa cccctgtggaa gatctgtggcc aaggtgcaggcc cccaggatccg caacaagaccc 840
 atctgtgttca agcagacgac cggggggggcc cccggatgttgc tgatgtccac gttcaactgc 900
 ggcggggagtt tctttactgtt caacacgacc cagctgttca acagacactg gaacaacacc 960
 atccggcccca aacaacacca cccggccatcc acctgtgttcc gccgcattca gcatcatac 1020
 aaccggggggcc gacccggccg catgtacccg ccccccactt gggcccgat cccgtgtggc 1080
 agcaacatca cccggccgttgc tgctggccgc gacggggccg aggagatccg caacaacacc 1140
 gagatcttcc gcccccggcc cggggacatgc cggcacaactt ggcgcacgcg gctgttacaag 1200
 tacaagggttgg tgaaatgttgc gcccctggcc tggtggccca ccaaggccaa ggcggccgtt 1260
 gtgcacggccg agaaaggccgc cgcggccatcc ggcggccatgt ttctgggtt cctggggcc 1320
 gccggccggca ccatggggcc cccggccatcc acctgtgttgc tgaggccccc ccggccgtt 1380
 agccggatccg tgccggccgcgaaacacccgttgcggccca tgccggccca gcaagccatcc 1440
 ctgcacgttgc cctgtgtgggg catcaaggccg ctgcaggcccc gctgtgttgc cgtggagccg 1500
 tacctgttgcagg acggcagatc gtcggccatc tggggttgc gggccaaat gatctgttcc 1560
 accccggccgttgc cttggaaacgcg cagctggccg aacaaggccg ttggccatgttgc 1620
 atgacccgttgc tggtggggcc ggcggccatcc gacaaatcaca ccaacccgttgc tcataccctgc 1680
 atcgaggaggcc gcaacacca gcaaggccgcgaaacccgttgc agtgcgttgc tgctggacaa 1740
 tggggccatcc tgggttgcgtt gttccatcc acgtgttgc tggtgtatcc caagatcttc 1800
 atcatgtatcc gtcggccgttgc tggtggccctt cgcacatgttgc tccacccgttgc gacccatgtgc 1860
 aaccggccgttgc gccacccggca cggccccccttgc acgttccca gccgtttccccc cggcccccggc 1920
 ggcggccggcc gcccccggccg catcgaggccg gaggggggccg agccggccacccg cggccggccgc 1980
 agcccccgttgc tgccacggccgttgc tggtggccctt atccggccatcc gctgtgttgc cctgtgttgc 2040
 ttctggccatcc accggccgttgc cgcactgttgc ctgatccggcc cccgcattgttgc ggtgtgttgc 2100
 ggcggccggccg gtcggggccg ctggaaatgttgc tggtggccaaatc tggtgttgc ctggatccca 2160
 gagctgttgcagg acacccggccgttgc gacccatccgcg cccatccggccgttgc gggccggccgc 2220
 accggccggca tcatcgagggttgc ggcggccatccgcg cttccctgttgc catcccccggc 2280
 cgcacatccggcc agggccctgttgc tggtgttgc ag 2322

<210> 20

<211> 2322

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Leu122-Ser199;
Trp427-Gly431

<400> 20

gaatccggccca ccatggatgc aatgaagaga gggctctgttgc tggtgtgttgc gctgtgtggaa 60
 gcacgtttccg ttttcggccac cgcggccgttgc aagctgtgtgg tgacccgttgc ctacggccgttgc 120
 cccgtgtggaa aggaggccac caccacccctt ttctggccca ggcacgcacca ggccttacgac 180
 accggagggtgc acacatgttgc ggcacccac gctgtgttgc ccacccgttgc caaccccccac 240
 gagatctgtc tggtggaaacgttgc gaccggaaat ttcacatgttgc gaaagacaa catgtgtggaa 300
 cagatgcacg aggacatcatcag ctggctgtgg gaccaggaccc tgaaggccctg cgtgaagctg 360
 gcaacacgcg tgcataccca ggcctgcccc aaggtgagct tcgagccat ccccatccac 420
 tactgtggcc cccggccgttgc cgcacatctgtg aagtgcacccg acaagaagtt caacggcagc 480
 ggccttgcac caacacgttgc caccgttgcac tgccatccggcc gatcccccggcc cgtgtgttgc 540
 accccggccgttgc tgctgaaacggc caccgttgcac tgccatccggcc gaggaggccg ttgtgtatccg 600
 ttccaccgaca acgccaacac catcatgttgc cagctgttgcaggaaagg agagctgttgc gatcaactgc 660

aacccggccca acaaacaac ccgcaagagc atcaccatcg gccccggccg cgccttctac 720
 gcccacggccg acatcatcg cgacatccgc cagggccact gcaacatcg cggcagaaga 780
 tggaaacaaca ccctgaagca gatctgtgacc aactgtcgagg cccagttcg caacaagacc 840
 atcgtgttca agcagagac cggccggcgc cccgagatcg tgatgcacag cttcaactcg 900
 ggcggcggagt tcttctactg caaacgacc cagctgttca acagacactt gaacaacacc 960
 atcggccccc acaaaccaaa cggcaccatc accctgcctt gccgcattca gtagatcatc 1020
 aaccgcgtgg ggcgcaaggc catgtacgc ccccccattc gggcccaatg cccgtcgacg 1080
 agcaacatca ccggccgtgt gctgacccgc gacggggca aggagatcg caacaccacc 1140
 gagatcttcc gcccggccgg cggcgcacatg cggcacaact ggccgcggca gctgtacaag 1200
 tacaaggtag tgaatgtca gcccctgggc gtggccccc ccaagggca ggcggcgtg 1260
 gttcggcggcg agaaggccgc cgtggacccgtt ggcgcgtatg tcttgggtt ccttggggc 1320
 gccggcggcga ccattggggc cccggccgtt accctgcgtt tcggcccg cccgtcgctg 1380
 agcggcgtcg tgacggcaga gaacaacccg ctgcgcggca tcgaggccca gcaacactcg 1440
 ctgcgtgtcc cctgtgggg catcaacggc ctgcgcggcc gctgtgtggc cttggagccg 1500
 tacctggagg acccaacgtt gctggggatc tggggcttc gggccaaatg gatctgcacc 1560
 accgcgtgc gcccggccgg cagctgtggc aacaaggacc ttggccaggat ctggaaacacc 1620
 atgaccctggg tggagtgggg ggcgcggatc gacaacttca ccaacccgtt ctacaccctg 1680
 atcggaggaga gcaaaacca gcaaggagaa aacggcaggc agtgcgttca gctggacaa 1740
 tggggccatcg tggggatcg agcaatgttc ttgggtatcat caagatcttc 1800
 atcatgtatcg tggggccgtt gctggggctt cccatgtgt tcacccgtgt gaggatcttg 1860
 aaccgcgtgc accggggatc cagggccctt agtgcgttca cccgttccc cggcccccggc 1920
 ggcggccgacc gcccggaggg catcgaggag gggggccgg aacggcggcc gcaacccggc 1980
 agcccccctgg tgcacggcctt gctggccctt atctggggac acctgtcgcc cctgtgtctg 2040
 ttcaatcgacc accggccgtt cgtatggccg cccgcattgtt ggactgtcg 2100
 ggcggccggccg tggggggggg ctggatgtt gggggcaacc ttgtcgatgtt ctggatccg 2160
 gagctgaaatc agacggccgtt gagctgttca gacgcatttc ccatcgccgtt ggccggggc 2220
 accggccgca tcatcgatgtt ggcccgccgc atccggccgg ctttctgtca catccccccg 2280
 cgcattccggc agggatctcg ggcggccctg ctgtaaatcg ag 2322

<210> 21

<211> 2310

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Lys121-Val1200;
Asn425-Lys432

<400> 21

gaatttcgcgc ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt gctgtgtgg 60
 gcaatgtttcg ttctggcccg cgcgtgtgg aagctgtggg tgaccgtgtt ctacggcg 120
 cccgtgtggg aggaggccgc caccatccgtt ttctggccca gggccaccaaa ggccatcagc 180
 accggagggtt aacaatgtgtt gggcccccac gctgtgtgg ccacggccacc caacccccc 240
 gagatctgtc tggaaatgtt gacggagacat ttcaatcgatg ggaaaacaaa catggttgg 300
 catatgcacg aggacatcatc cggctgtgg gacggacccgca tgaaggccctg ctgtggggcc 360
 cccgtgtatca cccaggctgtt ccccaagggtt agttcgaccc ccatccccat ccatactgtc 420
 gccccccgcgc gttccggatc ctggaaatgtt aacggacaaa agttcaacccgg cagggggccc 480
 tgcacccaaatcg tggacccatgtt gcaatgttcc cggccatgtt gggccgtgtt gaggccatgg 540
 ctgtgtgtca acggcggatc ggcggaggat ggggtgttca tccggcggca gaacttcacc 600
 gacaacgcgcgca agacccatcatc tggcgtatgtt aaggagatgg tggagatcaa ctgcaccccg 660
 cccaaacaca acacccggca gacatcatacc atcggccccc ggcggccctt ctacggccacc 720
 ggcgcacatca tggcgtatgtt cccggccggcc cactgttcaaa tcacggccggaa gaatggaa 780
 aacccctgtt agcagatgtt gaccaatgtt cggccggatgtt tcggcaacaa gaccatctgt 840
 ttcaacggatc gcaatgttcc gggccggccg cttccatgtt acatgttcaaa ctggccggcc 900
 gagttttttt actggacacatc caccatgtt ttcaacatgtt ccttggaaatca caccatcgcc 960
 cccaaacaca ccaacggccatc catccatgtt cccgtggccca tcaacatgtt catcaacggcc 1020
 cccaaaggccatc tggcgtatgtt cccggccggcc gggccggatgtt gtcggcggcc caacatcgacc 1080
 ggcggccgtgc tgacccggccg cggccggccgg gggccggatgtt gtcggcggcc acacccggccg 1140
 cccggccggccg gggccatgtt cggccggccgg tggatgttca acacccggccg gatcttccggc 1200

```

aaggatcgccg cccctggggct ggccccccacc aaggccaaagc gcccgtgtt gcagcgcgag 1260
aaggcgcccg tgacccttggg cggcatgttc ctggcttccggg gggccggccg cggccggacc 1320
ggccggccccc gggactggccat cggactggctt cggggccggcc agctgtctgg agccatggcc 1380
cagcggcaga aacaatcgctt gggccggccatc gggccggccggc agccatcgctt gcagctggacc 1440
ttgtggggca ttcggatcgcc gggccggccggc gtggccggccggc tggagggctt cttggaaaggc 1500
cagcggatgtc tggggatctgg gggccgtcggc ggcaatcgta tctggaccac cggccgtggcc 1560
tggaaacccca gttggggatcc caaaggatctt gggccatgtt ggaaacacat gacccgttggcc 1620
gatggggggc gggccatgttggcc caaactacccaa aacctgtatc acaccctgtt gggggggggcc 1680
cagaacccggc aggagaagaa cggccatgttggcc caaactacccaa aacctgtatc acaccctgtt gggggggggcc 1740
tggaaactgtt tgccatgttggcc caaactacccaa aacctgtatc acaccctgtt gggggggggcc 1800
ggccggccccc gggccggccccc gggccggccccc gggccggccccc gggccggccccc gggccggccccc 1860
ggccggccatca gggccggccatca gggccggccatca gggccggccatca gggccggccatca gggccggccatca 1920
ccggaggggca tggaggggcc gggccggccggcc gggccggccggcc acggccggccggcc tggaggggcc 1980
caccggccgtc tggggctgtt ctggggccggcc ctggccggccggcc tggggctgtt ctggccggcc 2040
ccggccggccggcc acggccggccggcc tggggctgtt ctggccggccggcc tggggctgtt ctggccggcc 2100
ttggggggccccc tggaaatctgg gggccaaatctgg ctggccgtactt gggatccggggcc tggggggccccc 2160
agccggccgtca gggccgttccgg cggccatcgcc atccggccggcc tggggggccccc tggggggccccc 2220
atccggatggcc cccggccatcgat tggggccggccccc tttccggccatcgat tggggggccccc tggggggccccc 2280
ggccggccgtcc gggccggccgtt ctggccggccccc tttccggccatcgat tggggggccccc tggggggccccc 2310

```

<210> 22

<211> 2298

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Val120-Ile201;
Ile424-Ala433

<400> 22

gacatcagca agtggctgtg gtatcatcaag atcttcatca tgatctgtggg cggctctgggt 1800
 ggccctcgccca tcgtgttca cgtgtgtggc acgtgttcaacc gctgtggccca gggctacacg 1860
 cccctgtact tcacaccggc cttcccccggg ccccgccggcc cccgacccggcc cggggccatc 1920
 gaggaggagg ggcggcgagcg cgacccggcagc cggcagcggcc cccctgtgtca gctaccacccg cttcgccgac 1980
 gcccctgtact gggacggactt ggcgcacccgt tgccctgttca gctaccacccg cttcgccgac 2040
 ctgtatctgtc tcggcgcccg catcggtggg ctgtgtggcc ggcggggctg ggaggccctg 2100
 aagtaactggg gcaactgtctg gcaactgtactgg atccaggagc tgaagaacacg cgcgtgagc 2160
 ctgttcacgcc ctaatcgccat cgcgtgtggcc gaggccacccg accgcataat cgaagggtggcc 2220
 cagcgatcg ggcgcgcctt cttgcacate cccgcgcgc tccgcgcaggg cttegagcgc 2280
 gcccctgtgt aactcgat 2298

<210> 23
 <211> 2298
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 Val1120-Ile201B; Ile424-Ala433

<400> 23
 gatattccca ccatggatc aatgaagaga gggctctgtc gtgtgtctgtc gctgtgtgg 60
 gcaatgttccg ttccggcccg cgcgcgtggg aagctgtggg tgaccgtgttca ctacggctgt 120
 cccctgtggc aggaggccggc caccacccgt ttcgtggccca gegacccaa ggcctacacg 180
 acccgagggtgc aacaatgtgtc ggcacccatc gctctgtgtc ccacccggcc caaccccccgg 240
 gagatctgtc tgccggacatcg gaccggacac ttcaactatgt ggaagaacatc catgggtggag 300
 cagatgcacg aggacatcatc caccgtgtgg gaccggaccc tgaagccctg cttggccggc 360
 atcaccacgg cttggcccca ggtgatgttc gggccatcc ccatcacta ctggggcccc 420
 gcccgggttc ctaatctgtca gtgcacccggc aagaatgttca acggcagggg cccctgtcacc 480
 aacgtgaccc cttgtgtggg caccacccggc atccggccgg tgggtggccat ccaatgtgtc 540
 ctgaaatggcc ggtggccggc ggaggccgtg tgatctgtcc ggcggaaactt caccggacac 600
 gcaaaacacca gttatgttca gcttggagg gacgtgtggc tcaactgtac cccggcccaac 660
 aacaacaccc gcaagacat caccatcgcc cccggccggc ctttctacgc caccggcgac 720
 atcattcgccg acatccggc gggccacttc aacatcgac gggagaatgtg gaaacacac 780
 ctggaaacggc tcgtgttccatc gttggccggc cttgtggccca acaagacatc cttgttcaag 840
 cagacggccg gggggggcccg cttgtgtggc atgcacccatgt tcaactgtggg cggcgaaatgtc 900
 ttctactgtca acacacccatc gttgttcaac acgttccatc gcaacacccat cggcccccac 960
 aacaccaacccg gcaatcatcacc cttgtccgtc cttgtggccggc agatcatcg cggcgccatg 1020
 taccggccccc cttatcgccg cttgtgtggc tcaactgtac acatccggc cttgtgtgtgg 1080
 accccggccacg gggccggggc gttatgttca caccacccatc ttttcggccg cggccggccgc 1140
 gacatcgccg acaactgtggg cttgtgtggc tacaatgtac aggtgtgtggatc gatcgagcc 1200
 ctggggccgg cccatccatc gggccatccatc cttgtgtgtggc acggcggagaa gggccggccgt 1260
 accctggccg cttatgttccatc gggccatccatc gggccggccg gcaatgttcaatc gggccggccgc 1320
 aacctgtggcc tggatgttccatc gggccggccg cttgtgtggc acgtatgtcc gcaatgtggcc 1380
 aacctgtggcc gggccatccatc gggccggccg cttgtgtggc aggtgtggcc tggatgttcaatc 1440
 aacgtatgtcc gggccggccggtt gttatgttccatc gggccggccg tggatgttccatc gggccggccgt 1500
 ggcattctggg gttatgttccatc gggccggccg tggatgttccatc gggccggccg tggatgttccatc 1560
 tggatgttccatc gggccggccg tggatgttccatc gggccggccg tggatgttccatc gggccggccg 1620
 gagatcgaccc actacacccatc cttgtgtggcc accctgtggcc aggtgtggcc gaaacccatcg 1680
 gagaatgtggcc acgtatgtccatc gggccggccg tggatgttccatc gggccggccg tggatgttccatc 1740
 gacatcgacca agtggctgtg tttatgttccatc tttatgttccatc gggccggccg tggatgttccatc 1800
 gggatgttccatc tttatgttccatc gggccggccg tggatgttccatc gggccggccg tggatgttccatc 1860
 cccctgtact tcacaccggc cttcccccggg cccggccggcc cttgtgtggcc tggatgttccatc 1920
 gaggaggagg ggcggcgaccc cttcccccggg cccggccggcc cttgtgtggcc tggatgttccatc 1980
 gcccctgtact gggacggactt ggcgcacccgt tttatgttccatc gttatgttccatc gggccggccg 2040
 ctgtatctgtc tcggcgcccg catcggtggg ctgtgtggcc gggccggccg tggatgttccatc 2100
 aagtaactggg gcaactgtctg gcaactgtactgg atccaggagc tgaagaacacg cgcgtgagc 2160
 ctgttcacgcc ctaatcgccat cttcccccggg cttcccccggg tggatgttccatc gggccggccg 2220
 cagcgatcg ggcgcgcctt cttgcacate cccgcgcgc tccgcgcaggg cttegagcgc 2280

<400> 25

<210> 26

<211> 2352

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Val127-Asn195,
Arg426-Gly431

<400> 26

gaatttccgc ccatggatgc aatgttggaaa gggcttgcgt gtgtgtgtgtc 60
gcaggatcttcg ttccgttcccg ccgttggggaa aagtgttgggg tgaccgggtta ctatccgggttca 120
ccctgttggaa aggaaatccgc caccatccatgtt tttccgttcccg gggacggatccaa ggttccatgtac 180
acccagggttcg acaaatccatgtt gggccatccatgtt ttcatgttcccg ccacccatgttccaa 240
gatagatgttc tgaaatccatgtt gggccatccatgtt ttcatgttcccg ccacccatgttccaa 300
catgtatccatgtt ccacatgttcccg gggccatccatgtt ttcatgttcccg ccacccatgttccaa 360
acccatgttcccg gggccatccatgtt ttcatgttcccg ccacccatgttccaa 420

